

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:29:30 ; Search time 60.99 Seconds
(without alignments)
119.849 Million cell updates/sec

Title: US-09-485-951-2
Perfect score: 1917
Sequence: 1 MAFSGSQAPYLPSPVPSCT.....LPTINLEVGDIQLTHVQT 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 197339 seqs, 20590346 residues

Total number of hits satisfying chosen parameters: 197339

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1633	85.2	311	3	US-08-946-914-4
2	783	40.8	149	2	US-08-788-584-3
3	661	34.5	145	2	US-08-788-584-1
4	603.5	31.5	324	3	US-08-946-914-11
5	574	29.9	323	1	US-08-469-667-16
6	574	29.9	323	3	US-08-946-914-2
7	574	29.9	323	3	PCT-US95-07289-16
8	535	27.9	145	2	US-08-788-584-5
9	535	27.9	145	3	US-08-946-914-12
10	472.5	24.6	316	4	US-09-131-648-5
11	471.5	24.6	317	3	US-08-946-914-6
12	454.5	23.7	316	2	US-08-728-521-3
13	454.5	23.7	316	2	US-08-647-960-2
14	454.5	23.7	316	3	US-08-946-914-15
15	454.5	23.7	316	3	US-08-946-914-17
16	353.5	18.4	264	2	US-08-728-521-1
17	328.5	17.1	264	1	US-08-562-311-4
18	321.5	16.8	336	4	US-09-131-648-1
19	321	16.7	262	3	US-08-946-914-14
20	318.5	16.6	250	3	US-08-946-914-10
21	317.5	16.6	250	1	US-08-562-311-2
22	255	13.3	200	3	US-08-946-914-8
23	254.5	13.3	177	2	US-08-647-960-6
24	250	13.0	147	2	US-08-647-960-7
25	230.5	12.0	136	3	US-08-946-914-13
26	226.5	11.8	135	2	US-08-647-960-5
27	198.5	10.4	132	2	US-08-647-960-11

28	189	9.9	131	2	US-08-647-960-9	Sequence 9, Appl
29	184	9.6	146	2	US-08-647-960-8	Sequence 8, Appl
30	179	9.3	184	2	US-08-647-960-10	Sequence 10, Appl
31	167	8.7	135	3	US-08-050-259B-18	Sequence 18, Appl
32	154	8.0	135	3	US-08-946-914-16	Sequence 16, Appl
33	154	8.0	135	3	US-08-050-259B-20	Sequence 20, Appl
34	154	8.0	135	4	US-09-489-292-2	Sequence 2, Appl
35	142.5	7.4	142	6	5242807-2	Patent No. 5242807
36	129	6.7	132	3	US-08-946-914-9	Sequence 9, Appl
37	115	6.0	402	6	5352575-5	Patent No. 5352575
38	112	5.8	135	2	US-08-647-960-3	Sequence 3, Appl
39	108.5	5.7	135	3	US-08-050-259B-16	Sequence 16, Appl
40	107	5.6	38	1	US-08-540-202-2	Sequence 2, Appl
41	105.5	5.5	366	6	5470718-4	Patent No. 5470718
42	96.5	5.0	511	1	US-08-220-151-17	Sequence 17, Appl
43	96.5	5.0	511	1	US-08-413-118-17	Sequence 17, Appl
44	96.5	5.0	511	3	US-08-473-446-17	Sequence 17, Appl
45	94.5	4.9	328	3	US-09-253-682-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1
US-08-946-914-4
; Sequence 4, Application US/08946914
; Patent No. 6027916
; GENERAL INFORMATION:
; APPLICANT: Ni, Jian
; APPLICANT: Gentz, Reiner L.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
; STREET: 1100 New York Ave., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946,914
; FILING DATE: Herewith
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/028,093
; FILING DATE: 09-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2500
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-946-914-4

Query Match 85.2%; Score 1633; DB 3; Length 311;
Best Local Similarity 87.6%; Pred. No. 3.2e-164;
Matches 311; Conservative 0; Mismatches 0; Indels 44; Gaps 1;
Oy 1 MAFSGSQAPYLPSPVPSCTIQGLDGLQITVNGTLVSSSCTREAVNFQTGFGNDIAF 60

Db 1 MAFSGSQAPYLSAVPFSGTIOGLDGLQITVNGTSSGTRFAVNFQTFSGNDIAF 60
QY 61 HFNPRFDGGYVVCNTRONGSWGPEERKTHMPFKGMPPDLCFLVSSDFKVMVNGTLFV 120
Db 61 HFNPRFDGGYVVCNTRONGSWGPEERKTHMPFKGMPPDLCFLVSSDFKVMVNGTLFV 120
QY 121 QYFHRVPFHRVDTSVNGSVOLSYISFQNPRTVPVQFAFSTVPFSQPVCFPPRGRRQK 180
Db 121 QYFHRVPFHRVDTSVNGSVOLSYISFQNPRTVPVQFAFSTVPFSQPVCFPPRGRRQK 180
QY 181 PGWTFANPAPITQTVIHTVQSAPGQMFSTPAIPPMYHPAYPMPFITTLILGLYPSKS 240
Db 149 -----TQTVIHTVQSAPGQMFSTPAIPPMYHPAYPMPFITTLILGLYPSKS 196
QY 241 ILLSGTVLPQAORFHNLCNHNIAFHLPNPRFDENAVVRNTQIDNSWGSEERSLPRKMPF 300
Db 197 ILLSGTVLPQAORFHNLCNHNIAFHLPNPRFDENAVVRNTQIDNSWGSEERSLPRKMPF 256
QY 301 VRQGSFVWILCEAHCLKVAVDGQHLFEYVHRLNRLPTINRLEVGGDIQLTHVQT 355
Db 257 VRQGSFVWILCEAHCLKVAVDGQHLFEYVHRLNRLPTINRLEVGGDIQLTHVQT 311

RESULT 2
US-08-788-584-3
; Sequence 3, Application US/08788584
; Patent No. 5837493
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Petithory, Joanne R.
; TITLE OF INVENTION: NOVEL HUMAN GALECTINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; FILING DATE: Filed Herewith
; APPLICATION NUMBER: 436
; PRIOR APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0192 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 149 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-788-584-3

Query Match 40.8%; Score 783; DB 2; Length 149;
Best Local Similarity 97.3%; Pred. No. 5.1e-75;

Matches 145; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 207 MFTPTAIPPMYHPAYPMPFITTLILGLYPSKILLSGTVLPQAORFHNLCNHNIAF 266
Db 1 MFTXGIPPMYHPGYPMPFITTLILGLYPSKILLSGTVLPQAORFHNLCNHNIAF 60
QY 267 HLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQHL 326
Db 61 HLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQHL 120
QY 327 FEYVHRLNRLPTINRLEVGGDIQLTHVQT 355
Db 121 FEYVHRLNRLPTINRLEVGGDIQLTHVQT 149

RESULT 3
US-08-788-584-1
; Sequence 1, Application US/08788584
; Patent No. 5837493
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Goli, Surya K.
; APPLICANT: Bandman, Olga
; APPLICANT: Hawkins, Phillip R.
; APPLICANT: Petithory, Joanne R.
; TITLE OF INVENTION: NOVEL HUMAN GALECTINS
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/788,584
; FILING DATE: Filed Herewith
; CLASSIFICATION: 436
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0192 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-788-584-1

Query Match 34.5%; Score 661; DB 2; Length 145;
Best Local Similarity 90.5%; Pred. No. 3.6e-62;
Matches 124; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 219 PHPAYPMPFITTLILGLYPSKILLSGTVLPQAORFHNLCNHNIAFHLPNPRFDENAV 278
Db 9 PYLSPXVPFSGTXQGLYPSKILLSGTVLPQAORFHNLCNHNIAFHLPNPRFDENAV 68
QY 279 RNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQHLFEYVHRLNRLPT 338
Db 69 RNNQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGQHLFEYVHRLNRLPT 128

QY 246 IVLP5AQRFHINLCSGN--HIAFHLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRG 303

[illegible]

Db 272 OPFDLSIRGCLDRFKVYANGOHLEFDFAHRLSAFORVDLTLEIQGDVTLVSQ 322

RESULT 8

US-08-788-584-5

; Sequence 5, Application US/08788584

; Patent No. 5837493

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Goli, Surya K.

; APPLICANT: Bandman, Olga

; APPLICANT: Hawkins, Phillip R.

; APPLICANT: Pettithory, Joanne R.

; TITLE OF INVENTION: NOVEL HUMAN GALECTINS

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Drive

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/788,584

; FILING DATE: Filed Herewith

; CLASSIFICATION: 436

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0192 US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-855-0555

; TELEFAX: 415-845-4166

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 145 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 727176

US-08-788-584-5

Query Match 27.9%; Score 535; DB 2; Length 145;

Best Local Similarity 70.3%; Pred. No. 7.2e-49;

Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;

QY 208 FSTPAIPPMYPHPAYMPFITTILGGLYPSKILLSGTVLPSAQRHINLCSGNHIAFH 267

Db 4 FSTQT-----PYPNLAVPFTSIPNGLYPSKISIVISGVVLSDAKRFQINLRCGGDIAFH 57

QY 268 LNPREDENAVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGOHLF 327

Db 58 LNPREDENAVRNTQINNSWGPEERSLPGSMFPSRGQSFVWILCEHGCFCFKVAVDGOHC 117

QY 328 EYHRLRLNLPNTINRLEVGDDIQLTHVQT 355

Db 118 EYSHRLMLNLPDINTLEVAGDIQLTHVET 145

RESULT 9

US-08-946-914-12

; Sequence 12, Application US/08946914

; Patent No. 6027916

; GENERAL INFORMATION:

; APPLICANT: Ni, Jian

; APPLICANT: Gentz, Reiner L.

; APPLICANT: Ruben, Steven M.

; TITLE OF INVENTION: Galectin 8, 9, 10 and 10SV

; NUMBER OF SEQUENCES: 60

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.

; STREET: 1100 New York Ave., Suite 600

; CITY: Washington

; STATE: D.C.

; COUNTRY: USA

; ZIP: 20005-3934

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/946,914

; FILING DATE: Herewith

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/028,093

; FILING DATE: 09-OCT-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Steffe, Eric K.

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 1488.0560001/EKS/SGW

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2540

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 145 amino acids

; TYPE: amino acid

; STRANDEDNESS: not relevant

; TOPOLOGY: linear

; MOLECULE TYPE: CDNA

US-08-946-914-12

Query Match 27.9%; Score 535; DB 3; Length 145;

Best Local Similarity 70.3%; Pred. No. 7.2e-49;

Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;

QY 208 FSTPAIPPMYPHPAYMPFITTILGGLYPSKILLSGTVLPSAQRHINLCSGNHIAFH 267

Db 4 FSTQT-----PYPNLAVPFTSIPNGLYPSKISIVISGVVLSDAKRFQINLRCGGDIAFH 57

QY 268 LNPREDENAVRNTQIDNSWGSEERSLPRKMPFVRGQSFVWILCEAHCLKVAVDGOHLF 327

Db 58 LNPREDENAVRNTQINNSWGPEERSLPGSMFPSRGQSFVWILCEHGCFCFKVAVDGOHC 117

QY 328 EYHRLRLNLPNTINRLEVGDDIQLTHVQT 355

Db 118 EYSHRLMLNLPDINTLEVAGDIQLTHVET 145

RESULT 10

US-09-131-648-5

; Sequence 5, Application US/09131648

; Patent No. 6168920

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Yue, Henry

; APPLICANT: Corley, Neil C.

; APPLICANT: Guegler, Karl J.

; APPLICANT: Patterson, Chandra

; TITLE OF INVENTION: EXTRACELLULAR ADHESIVE PROTEINS

; FILE REFERENCE: PF-0576 US

RESULT 12
US-08-728-521-3
Sequence 3, Application US/08728521
Patent No. 5869289
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: HUMAN GALECTIN HOMOLOG
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,521
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0137 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 717032
US-08-728-521-3

Query Match 23.7%; Score 454.5; DB 2; Length 316;
Best Local Similarity 31.7%; Pred. No. 7.3e-40;
Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps 8;

QY 1 MAFSGSQAPYLPSPAFSGTIOGGLODGLQITVNGTVLSSSGTREFAVNFOTGFS---GND 57
DB 2 LSLSLQNIYIYPTIYVSTITEQLKPGSLIVIRGHVPKDS-ERFQVDFQHGNSLKPRAD 60
QY 58 IAFHENPRFEDGGYVVCNTRQNGSWGPERKTHMPFQKGMPEFLCFLVQSSDFKVMVNGI 117
DB 61 VAFHENPREKRSNCIVCNTLTNEKGWEEITHDMFERKESEIYIMVLKKNFHVAVNGK 120
QY 118 LFVQYFHRVPHRVDTISVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRPRGR 177
DB 121 HLLVAHRINPEKIDTLGIFGKVNHSIGR-----FSSDLQSM----- 159
QY 178 RQKPGVWPANPAPTQVIHTVQSAQGMFSTPAIPPMYPHPAYPMPFITTLGGLYP 237
DB 160 -----ETSLGLTQISKENIQKS-GKLHLS-----LPFEARLNASMG 196
QY 238 SKSILLSTGLVPSAQRHINLCSG--NHIAFLNPRFDENAVVRNTQIDNSWGSEERSLP 295
DB 197 GRTVVVKGEVNTNATSFNVDLVAGRSRDIALHLNPLNKAFAVRNSFLQDANGEERNI- 255
QY 296 RKMFPVQSGFSWILCEAHCLKVAVDGOHLFEYHRLNRLPTINRLEVGDIQTHVOT 355
DB 256 TCFPFSSGMFYEMIIYCDVREFKAVNGVHSLYKHREKDLSSIDTLAVDGDGIRLLDVS 315

RESULT 13
US-08-647-960-2
Sequence 2, Application US/08647960
Patent No. 5908761
GENERAL INFORMATION:
APPLICANT: ZICK, Yehiel
TITLE OF INVENTION: GALECTIN-8 AND GALECTIN-8-LIKE PROTEINS
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,960
FILING DATE: 30-MAY-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: IL 107880
FILING DATE: 05-DEC-1993

ATTORNEY/AGENT INFORMATION:
NAME: BROWDY, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: ZICK-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3527
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 316 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-647-960-2

Query Match 23.7%; Score 454.5; DB 2; Length 316;
Best Local Similarity 31.7%; Pred. No. 7.3e-40;
Matches 114; Conservative 62; Mismatches 133; Indels 51; Gaps 8;

QY 1 MAFSGSQAPYLPSPAFSGTIOGGLODGLQITVNGTVLSSSGTREFAVNFOTGFS---GND 57
DB 2 LSLSLQNIYIYPTIYVSTITEQLKPGSLIVIRGHVPKDS-ERFQVDFQHGNSLKPRAD 60
QY 58 IAFHENPRFEDGGYVVCNTRQNGSWGPERKTHMPFQKGMPEFLCFLVQSSDFKVMVNGI 117
DB 61 VAFHENPREKRSNCIVCNTLTNEKGWEEITHDMFERKESEIYIMVLKKNFHVAVNGK 120
QY 118 LFVQYFHRVPHRVDTISVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRPRGR 177
DB 121 HLLVAHRINPEKIDTLGIFGKVNHSIGR-----FSSDLQSM----- 159
QY 178 RQKPGVWPANPAPTQVIHTVQSAQGMFSTPAIPPMYPHPAYPMPFITTLGGLYP 237
DB 160 -----ETSLGLTQISKENIQKS-GKLHLS-----LPFEARLNASMG 196
QY 238 SKSILLSTGLVPSAQRHINLCSG--NHIAFLNPRFDENAVVRNTQIDNSWGSEERSLP 295
DB 197 GRTVVVKGEVNTNATSFNVDLVAGRSRDIALHLNPLNKAFAVRNSFLQDANGEERNI- 255
QY 296 RKMFPVQSGFSWILCEAHCLKVAVDGOHLFEYHRLNRLPTINRLEVGDIQTHVOT 355
DB 256 TCFPFSSGMFYEMIIYCDVREFKAVNGVHSLYKHREKDLSSIDTLAVDGDGIRLLDVS 315

RESULT 14
US-08-946-914-15
Sequence 15, Application US/08946914
Patent No. 6027916
GENERAL INFORMATION:
APPLICANT: NI, Jian
APPLICANT: Gentz, Reiner L.
APPLICANT: Ruben, Steven M.
TITLE OF INVENTION: Galectin 8, 9, 10 and 10S9
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein, & Fox P.L.L.C.
STREET: 1100 New York Ave., Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,914
FILING DATE: Herewith
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,093

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:42:18 ; Search time 36.57 Seconds
(without alignments)
332.532 Million cell updates/sec

Title: US-09-485-951-2

Perfect score: 1917

Sequence: 1 MAFSGSQAPYLPSPVFSCT.....LPTINRLEVGGDIQLTHVQT 355

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1917	100.0	355	LEG9_HUMAN	O00182 homo sapien
2	1374.5	71.7	354	LEG9_RAT	P07840 rattus norv
3	1324	69.1	353	LEG9_MOUSE	O08573 mus musculus
4	603.5	31.5	324	LEG4_RAT	P38552 rattus norv
5	574	29.9	323	LEG4_HUMAN	P56470 homo sapien
6	541	28.2	323	LEG4_PIG	Q29058 sus scrofa
7	535	27.9	144	LEG5_RAT	P47967 rattus norv
8	518	27.0	301	LEG6_MOUSE	O34891 mus musculus
9	467.5	24.4	316	LEG6_MOUSE	Q91115 mus musculus
10	463.5	24.2	316	LEG8_HUMAN	O00214 homo sapien
11	454.5	23.7	316	LEG8_RAT	Q62665 rattus norv
12	427	22.3	283	LEG1_HAECO	O44126 haemochus
13	414.5	21.6	279	LEG3_CABEL	P36573 caenorhabdi
14	374	19.5	285	LEG3_CABEL	Q09581 caenorhabdi
15	339	17.7	244	LEG3_CRITLO	P47953 cricetus
16	330.5	17.2	295	LEG3_CANFA	P38486 canis famil
17	324.5	16.9	241	LEG3_RABIT	P47845 oryctolagus
18	322.5	16.8	263	LEG3_MOUSE	P16110 mus musculus
19	321	16.7	261	LEG3_RAT	P08699 rattus norv
20	318.5	16.0	239	LEG3_HUMAN	P17931 homo sapien
21	230.5	12.0	135	LEG7_HUMAN	P47929 homo sapien
22	223.5	11.7	135	LEG7_MOUSE	O54974 mus musculus
23	216.5	11.3	135	LEG7_RAT	P07590 rattus norv
24	189.5	9.9	134	LEG6_CHICK	P32668 gallus gall
25	184.5	9.6	134	LEG4_CHICK	P07583 gallus gall
26	182	9.5	182	LEG7_CABEL	Q09605 caenorhabdi
27	167	8.7	134	LEG1_MOUSE	P16045 mus musculus
28	166	8.7	134	LEG1_SHEEP	P81184 ovis aries
29	164	8.6	134	LEG1_CRIGR	P48538 cricetus
30	164	8.6	134	LEG1_RAT	P11762 rattus norv
31	164	8.6	184	LEG8_CABEL	Q09610 caenorhabdi
32	160	8.3	134	LEG1_BOVIN	P11116 bos taurus
33	155	8.1	129	LEG_ELEEL	P08520 electrophor

RESULT 1

LEG9_HUMAN	154	8.0	134	1	LEG1_HUMAN	P09382 homo sapien
ID	LEG9_HUMAN	STANDARD;	PRT;	355 AA.		Q9YIC2 conger myri
AC	O00182; O14532; O75028;					Q05315 homo sapien
DT	01-NOV-1997 (Rel. 35, Created)					Q92144 rattus norv
DT	15-JUL-1998 (Rel. 36, Last sequence update)					P11946 mus musculus
DT	01-OCT-2000 (Rel. 40, Last annotation update)					P05162 homo sapien
DE	GALECTIN-9 (HOM-HD-21) (ECALECTIN).					P36217 bufo arenar
GN	LGALS9.					Q09277 caenorhabdi
OS	Homo sapiens (Human).					P07645 pseudorhabdi
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					P26788 conger myri
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.					Q09606 caenorhabdi
OX	NCBI_TaxID=9606;					
[1]						
RN	SEQUENCE FROM N.A. (SHORT FORM).					
RP	TISSUE=Splice;					
RC	MEDLINE=97197815; PubMed=9045665;					
RA	Tuercel O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;					
RT	"Molecular definition of a novel human galectin which is immunogenic					
RT	in patients with Hodgkin's disease.";					
RL	J. Biol. Chem. 272:6416-6422(1997).					
[2]						
RP	SEQUENCE FROM N.A. (LONG FORM).					
RC	TISSUE=Gastric carcinoma;					
RA	Kato S.;					
RN	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.					
[3]						
RP	SEQUENCE FROM N.A. (SHORT FORM).					
RC	MEDLINE=98307937; PubMed=9642261;					
RA	Matsumoto R., Matsumoto H., Seki M., Hata M., Asano Y., Kanegasaki S.,					
RA	Stevens R.L., Hirashima M.;					
RT	"Human ealectin, a variant of human galectin-9, is a novel eosinophil					
RT	chemoattractant produced by T lymphocytes.";					
RL	J. Biol. Chem. 273:16976-16984(1998).					
CC	- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-					
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE					
CC	SHORT ISOFORM ACTS AS AN EOSINOPHIL-CHEMOATTRACTANT.					
CC	- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN					
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.					
CC	- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC					
CC	TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.					
CC	- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING					
CC	DOMAINS.					
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.					
CC	-----					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@isb-sib.ch)					
CC	-----					
CC	EMBL; Z49107; CAA8922.1; -					
DR	EMBL; AB006782; BAA22166.1; -					
DR	EMBL; AB005894; BAA31542.1; -					

DR HSSP; P17931; 1A3K.
 DR MIM; 601879; -
 DR InterPro; IPR001079; -
 DR Pfam; PF00337; Gal-bind-lectin; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 DR Galaptin; Lectin; Repeat; Alternative splicing.
 KW DOMAIN 1 148 GALAPTIN 1.
 FT DOMAIN 1 148 LINKER.
 FT DOMAIN 149 205
 FT DOMAIN 207 355
 FT BINDING 82 88 BETA-GALACTOSIDE (BY SIMILARITY).
 FT BINDING 88 88 BETA-GALACTOSIDE (BY SIMILARITY).
 FT BINDING 287 293 MISSING (IN SHORT ISOFORM).
 FT VARSPLIC 149 180
 FT CONFLICT 5 5 G -> S (IN REF. 3).
 FT CONFLICT 88 88 K -> R (IN REF. 1).
 FT CONFLICT 135 135 S -> F (IN REF. 1).
 FT CONFLICT 270 270 P -> L (IN REF. 1).
 FT CONFLICT 313 313 E -> G (IN REF. 1).
 SQ SEQUENCE 355 AA; 39518 MW; 4748C222FCAFA536A CRC64;

Query Match 100.0%; Score 1917; DB 1; Length 355;
 Best Local Similarity 100.0%; Pred. No. 1.3e-146;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFSGSQAPYLSPAVPFSGTIQGLQDGLQITVNGTVLSSGTRFAVNFOTGFGSGNDIAF 60
 DB 1 MAFSGSQAPYLSPAVPFSGTIQGLQDGLQITVNGTVLSSGTRFAVNFOTGFGSGNDIAF 60
 QY 61 HFNPRFEDGGYVVCNTRQNGSWGPEERKTHMPFQKGMPPDCLFLVSSDFKVMVNGILFV 120
 DB 61 HFNPRFEDGGYVVCNTRQNGSWGPEERKTHMPFQKGMPPDCLFLVSSDFKVMVNGILFV 120
 QY 121 QYFHRVPFHRVDTISVNGSVQLSYISFQNPRTVPQPAFSTVPFQVCPFPFRPRGRROK 180
 DB 121 QYFHRVPFHRVDTISVNGSVQLSYISFQNPRTVPQPAFSTVPFQVCPFPFRPRGRROK 180
 QY 181 PGVWPANPAPITQTVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITITLGLGYPSKS 240
 DB 181 PGVWPANPAPITQTVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITITLGLGYPSKS 240
 QY 241 ILLSGTVLPSAQRFHNLCSGNHIAFHLNPRFDENAVVNTQIDNSWGSEERSLPRKMPF 300
 DB 241 ILLSGTVLPSAQRFHNLCSGNHIAFHLNPRFDENAVVNTQIDNSWGSEERSLPRKMPF 300
 QY 301 VRQGSFVSWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDIQLTHVQT 355
 DB 301 VRQGSFVSWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDIQLTHVQT 355

SULT 2
 9_RAT
 AC LEG9_RAT STANDARD; PRT; 354 AA.
 DT P97840; O08588; O35866;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GALECTIN-9 (36 KDA BETA-GALACTOSIDE BINDING LECTIN) (URATE
 DE TRANSPORTER/CHANNEL) (UAT).
 GN LGALS9.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Kidney; and Small intestine;
 RX MEDLINE=97150351; PubMed=9038233;
 RA Wada J., Kanwar Y.S.;
 RT "Identification and characterization of galectin-9, a novel beta-
 RT galactoside-binding mammalian lectin.";
 RL J. Biol. Chem. 272:6078-6086(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (SHORT FORM).

RC STRAIN-SPRAGUE-DAWLEY; TISSUE-Kidney;
 RX MEDLINE=97150769; PubMed=8995305;
 RA Leal-Pinto E., Tao W., Rappaport J., Richardson M., Knorr B.A.,
 RA Abramson R.G.;
 RT "Molecular cloning and functional reconstitution of a urate
 RT transporter/channel.";
 RL J. Biol. Chem. 272:617-625(1997).
 CC -!- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-
 CC EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS (BY
 CC SIMILARITY). MAY PROVIDE THE MOLECULAR BASIS FOR URATE FLUX ACROSS
 CC CELL MEMBRANES, ALLOWING URATE THAT IS FORMED DURING PURINE
 CC METABOLISM TO EFFLUX FROM CELLS AND SERVING AS AN ELECTROGENIC
 CC TRANSPORTER THAT PLAYS AN IMPORTANT ROLE IN RENAL AND
 CC GASTROINTESTINAL URATE EXCRETION. HIGHLY SELECTIVE TO THE ANION
 CC URATE.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-
 CC CLASSICAL SECRETORY PATHWAY (BY SIMILARITY).
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN
 CC HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE
 CC SMALL INTESTINE.
 CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
 CC DOMAINS.
 CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U59462; AAB51192.1; -
 DR EMBL; U72741; AAB68592.1; -
 DR EMBL; U67958; AAB48591.1; -
 DR HSSP; P17931; 1A3K.
 DR InterPro; IPR001079; -
 DR Pfam; PF00337; Gal-bind-lectin; 2.
 DR PROSITE; PS00309; GALAPTIN; 2.
 DR Galaptin; Lectin; Repeat; Alternative splicing; Ion transport.
 KW DOMAIN 1 147 GALAPTIN 1.
 FT DOMAIN 1 147 LINKER.
 FT DOMAIN 148 205
 FT DOMAIN 206 354 GALAPTIN 2.
 FT BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).
 FT BINDING 286 292 BETA-GALACTOSIDE (BY SIMILARITY).
 FT VARSPLIC 148 179 MISSING (IN SHORT ISOFORM).
 SQ SEQUENCE 354 AA; 39946 MW; 6574F960B2EAF37C CRC64;

Query Match 71.7%; Score 1374.5; DB 1; Length 354;
 Best Local Similarity 71.8%; Pred. No. 3.7e-103;
 Matches 255; Conservative 31; Mismatches 68; Indels 1; Gaps 1;

QY 1 MAFSGSQAPYLSPAVPFSGTIQGLQDGLQITVNGTVLSSGTRFAVNFOTGFGSGNDIAF 60
 DB 1 MAFSGSQAPYLSPAVPFSGTIQGLQDGLQITVNGTVLSSGTRFAVNFOTGFGSGNDIAF 60
 QY 61 HFNPRFEDGGYVVCNTRQNGSWGPEERKTHMPFQKGMPPDCLFLVSSDFKVMVNGILFV 120
 DB 61 HFNPRFEDGGYVVCNTRQNGSWGPEERKTHMPFQKGMPPDCLFLVSSDFKVMVNGILFV 120
 QY 121 QYFHRVPFHRVDTISVNGSVQLSYISFQNPRTVPQPAFSTVPFQVCPFPFRPRGRROK 180
 DB 121 QYFHRVPFHRVDTISVNGSVQLSYISFQNPRTVPQPAFSTVPFQVCPFPFRPRGRROK 180
 QY 181 PGVWPANPAPITQTVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITITLGLGYPSKS 240
 DB 181 PGVWPANPAPITQTVIHTVQSPAGQMFSTPAIPPMYPHPAYPMPFITITLGLGYPSKS 240
 QY 241 ILLSGTVLPSAQRFHNLCSGNHIAFHLNPRFDENAVVNTQIDNSWGSEERSLPRKMPF 300
 DB 241 ILLSGTVLPSAQRFHNLCSGNHIAFHLNPRFDENAVVNTQIDNSWGSEERSLPRKMPF 300
 QY 301 VRQGSFVSWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDIQLTHVQT 355
 DB 301 VRQGSFVSWLCEAHLCKVAVDQHLFEYHRLNRLPTINRLEVGDIQLTHVQT 355

QY	301	VRQSF	SVWILCEAHCLKVAVDQGHLEFYHRLRNLP	TINRLEVGGDIQLTHVQT	355
DB	300	NRQGSF	SVWILCEBCHGCFKAVDQGHICEYHRLKNLP	DINTLEVAGDIQLTHVQT	354
RESULT	3				
LEG9_MOUSE					
ID	LEG9_MOUSE	STANDARD;	PRT;	353	AA.
AC	008573;	008572;			
DT	01-NOV-1997	(Rel. 35, Created)			
DT	01-NOV-1997	(Rel. 35, Last sequence update)			
DT	01-OCT-2000	(Rel. 40, Last annotation update)			
DE	GALECTIN-9.				
GN	LGALS9.				
OS	Mus musculus (Mouse).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	NCBI_TaxID=10090;				
	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CD-1; TISSUE=Small intestine, and Kidney;				
RX	MEDLINE=97190351; PubMed=9038233;				
RA	Wada J., Kanwar Y.S.;				
RA	"Identification and characterization of galectin-9, a novel beta-				
RT	galactoside-binding mammalian lectin.";				
RL	J. Biol. Chem. 272:6078-6086(1997).				
RN	[2]				
RP	CHARACTERIZATION.				
RX	MEDLINE=97298141; PubMed=9153289;				
RA	Wada J., Ota K., Kumar A., Wallner E.I., Kanwar Y.S.;				
RA	"Developmental regulation, expression, and apoptotic potential of				
RT	galectin-9, a beta-galactoside binding lectin.";				
RL	J. Clin. Invest. 99:2452-2461(1997).				
CC	-1- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-				
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS.				
CC	-1- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY ALSO BE SECRETED BY A NON-				
CC	CLASSICAL SECRETORY PATHWAY.				
CC	-1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN				
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	-1- TISSUE SPECIFICITY: ACCENTUATED EXPRESSION IN LIVER AND THYMUS OF				
CC	EMBRYO, DETECTED IN EMBRYONIC HEART, BRAIN, LUNG, LIVER, AND				
CC	KIDNEY. HIGHLY EXPRESSED IN ADULT THYMUS, SMALL INTESTINE, AND				
CC	LIVER, AND TO A LESSER EXTENT IN LUNG, KIDNEY, SPLEEN, CARDIAC,				
CC	AND SKELETAL MUSCLE. BARELY DETECTABLE IN BRAIN AND RETICULOCYTE.				
CC	THE LONG FORM IS EXPRESSED EXCLUSIVELY IN THE SMALL INTESTINE.				
CC	-1- DEVELOPMENTAL STAGE: THE EXPRESSION INCREASED WITH SUCCESSIVE				
CC	STAGES OF EMBRYONIC DEVELOPMENT.				
CC	-1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING				
CC	DOMAINS.				
CC	-1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U55061; AAB51190.1; -				
DR	EMBL; U55060; AAB51189.1; -				
DR	HSSP; P17931; 1A3K.				
DR	MGI; MGI:109496; Lgals9.				
DR	InterPro; IPR001079; -				
DR	Fram; PF00337; Gal_bind_lectin; 2.				
DR	PROSITE; PS00309; GALAPTIN; 2.				
KW	Galaptin; Lectin; Repeat; Alternative splicing.				
FT	DOMAIN 1 147 GALAPTIN 1.				
FT	FT 148 204 LINKER.				
FT	DOMAIN 205 353 GALAPTIN 2.				
FT	BINDING 81 87 BETA-GALACTOSIDE (BY SIMILARITY).				
FT	BINDING 285 291 BETA-GALACTOSIDE (BY SIMILARITY).				

```

-----
CC EMBL; M73553; AAA41505.1; -.
DR PIR; A46631; A46631.
DR HSP; PI7931; IA3K.
DR InterPro; IPR001079; -.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152
FT DOMAIN 153 177
FT DOMAIN 178 324
FT BINDING 257 263
FT BINDING 257 263
SQ SEQUENCE 324 AA; 36347 MW; 478024D7322AFE7B CRC64;

Query Match 31.5%; Score 603.5; DB 1; Length 324;
Best Local Similarity 39.6%; Pred. No. 2.le-41;
Matches 139; Conservative 54; Mismatches 115; Indels 43; Gaps 10;

QY 9 PYLSPAVPFGTIOGGLQDGLQITVNGTVLSSGTRFAVNFQGTG-FSGNDIAFHFNPRE 67
11 PTYNPLPKRPIPGSLVGMSSYIQG-TAKDMRRHFVNFAVGQDEGADIAFHFNPRE 69
68 DGGYVVCNTRQSGWGPPEERKTHMPFQKMPDCLFVQSSDFKVMVNGILFVQYFHRVP 127
70 GWDKVVNTMQSGWCKEKKSMPEKQGHFELVPMVMSHYKVVVNGTPEYEGHRLP 129
128 FHRVDIISVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRRGRKQKPGVWPA 187
130 LQMVTHLVQDGLQGLQINF-----LGGQFAASQYPTMTI-----PA 167
188 NPAPITQTVIHTVQSGAPGQMFSTPAI--PPMMYPHPAYPMPFTITLGGLYPSKILLSG 245
168 YP-----SAGYNPPQMSLVPWAGPIFNP-----PPIVGTGGLTARRTIIGK 214
246 TVLPSAQRHINLCGN--HIAFLNPRFDENAVVNTQIDNSGSEERSLPRKMPFVRG 303
215 YVLTAKLIINFKVSGTGDIAPHMNPRIQD-CVVRNSYMWNGSWGSEERKIPYN-PFGAG 272
304 QSFVSWILCEAHLKVAVDGOHLFEYFHHRLNPLTNRLNLEVGDDIOLTHVQ 354
273 QFFDLISRCGTRDFKVFANGQHLFDFSHRFQAFQVRVDMLEIKGDIPLSYVQ 323

RESULT 5
LEG4_HUMAN
ID LEG4_HUMAN STANDARD; PRT; 323 AA.
AC P56470;
DT 15-JUL-1998 (Rel. 36, Created)
15-JUL-1998 (Rel. 36, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
DE (L36LBP).
GN LGALS4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Gastric carcinoma;
RA Kato S.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colorectal carcinoma;
RA Rechreche H., Mallo G.V., Montalto G., Dagorn J.C., Iovanna J.L.;
RX MEDLINE=97454307; PubMed=9310382;
RT "Cloning and expression of the mRNA of human galectin-4, an S-type
RT lectin down-regulated in colorectal cancer.";
RL Eur. J. Biochem. 248:225-230(1997).
CC -!- FUNCTION: GALECTIN THAT BINDS LACTOSE AND A RELATED RANGE OF
CC SUGARS. MAY BE INVOLVED IN THE ASSEMBLY OF ADHERENS JUNCTIONS.
-----

```

```

-----
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
CC DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB006781; BAA22165.1; -.
DR EMBL; AF014838; AAC51763.1; -.
DR HSP; PI7931; IA3K.
DR MIM; 602518; -.
DR InterPro; IPR001079; -.
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR PROSITE; PS00309; GALAPTIN; 2.
KW Galaptin; Lectin; Repeat.
FT DOMAIN 1 152
FT DOMAIN 153 177
FT DOMAIN 178 323
FT BINDING 256 262
FT BINDING 256 262
SQ SEQUENCE 323 AA; 35941 MW; E79BC0A9AB3990EF CRC64;

Query Match 29.9%; Score 574; DB 1; Length 323;
Best Local Similarity 38.2%; Pred. No. 4.8e-39;
Matches 134; Conservative 52; Mismatches 121; Indels 44; Gaps 10;

QY 9 PYLSPAVPFGTIOGGLQDGLQITVNGTVLSSGTRFAVNFQGTGFS-GNDIAFHFNPRE 67
11 PTYNPLPKRPIPGSLVGMSSYIQG-VASEHMKRFFNVVVGQDGSVAFHFNPRE 69
68 DGGYVVCNTRQSGWGPPEERKTHMPFQKMPDCLFVQSSDFKVMVNGILFVQYFHRVP 127
70 GWDKVVNTMQSGWCKEKKSMPEKQGHFELVPMVMSHYKVVVNGTPEYEGHRLP 129
128 FHRVDIISVNGSVQLSYISFQNPRTVPVQPAFSTVPFSPVCFPPRRGRKQKPGVWPA 187
130 LQMVTHLVQDGLQGLQINFQGG-----PLRPQG-----PPMMP 165
188 NPAPITQTVIHTVQSGAPGQMFSTPAI--PPMMYPHPAYPMPFTITLGGLYPSKILLSG 245
166 YP-----GHCHQ-----QLNSLPTMEGPTFNP-----PVPYFGRLLQGLTARRTIIGK 212
246 TVLPSAQRHIN--LCSGNHIAFLNPRFDENAVVNTQIDNSGSEERSLPRKMPFVRG 303
213 YVPTGKSFAINFKVSGSGDIALHINPRMGNGTVVRNSLNGSWGSEERKITHN-PFGPG 271
304 QSFVSWILCEAHLKVAVDGOHLFEYFHHRLNPLTNRLNLEVGDDIOLTHVQ 354
272 QFFDLISRCGLDRFKVYANGQHLFDFHRLSAFQVRVDTLEIQDGVTLSYVQ 322

RESULT 6
LEG4_PIG
ID LEG4_PIG STANDARD; PRT; 323 AA.
AC Q29058; Q29296;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE GALECTIN-4 (LACTOSE-BINDING LECTIN 4) (L-36 LACTOSE BINDING PROTEIN)
DE (L36LBP).
GN LGALS4.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
-----

```

QY	353 VQ 354
DB	321 VQ 322
RESULT 7	
LEG5_RAT	
ID	LEG5_RAT
AC	P47967;
DT	01-FEB-1996 (Rel. 33, Created)
DT	01-FEB-1996 (Rel. 33, Last sequence update)
DT	01-FEB-1996 (Rel. 33, Last annotation update)
DE	GALECTIN-5 (RL-18).
GN	LGALS5.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON	NCBI_TaxID=10116;
OX	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Reticulocytes;
RC	MEDLINE=95197487; PubMed=7890611;
RA	Gitt M.A., Wiser M.F., Leffler H., Herrmann J., Xia Y.-R.,
RA	Massa S.M., Cooper D.N.W., Lusis A.J., Barondes S.H.;
RT	"Sequence and mapping of galectin-5, a beta-galactoside-binding
RT	lectin, found in rat erythrocytes";
RL	J. Biol. Chem. 270:5032-5038(1995).
CC	-!- FUNCTION: MAY FUNCTION IN ERYTHROCYTE DIFFERENTIATION.
CC	-!- SUBUNIT: MONOMER.
CC	-!- TISSUE SPECIFICITY: ERYTHROCYTES.
CC	-!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L36862; AAC42050.1; -
DR	EMBL; L21711; AAA65445.1; -
DR	HSP; P17931; IAKK
DR	InterPro: IPR001079; -
DR	Fram; PF00337; Gal-bind_lectin; 1.
DR	PROSITE; PS00309; GALAPTIN; 1.
DR	KW Galaptin; Lectin; Acetylation.
FT	INIT_MET 0 0
FT	MOD_RES 1 1
FT	FT BINDING 76 82
FT	CONFLICT 127 135
FT	SEQUENCE 144 AA; 16065 MW; BC95283D760DA515 CRC64;
FT	-----
FT	ACETYLATION.
FT	BETA-GALACTOSIDE (POTENTIAL).
FT	MISSING (IN AAA65445).
FT	-----
FT	Query Match 27.9%; Score 535; DB 1; Length 144;
FT	Best Local Similarity 70.3%; Pred. No. 2.5e-36;
FT	Matches 104; Conservative 9; Mismatches 29; Indels 6; Gaps 1;
FT	-----
QY	208 FSTPAIPMMYHPAYPMPTITTLGLYPSKSLISGTVLPSAQRPHINLCSGNHIAFH 267
DB	: : : : : : : : : : : : : : :
DB	3 FSTQ-----PYPNLAVPFTSPINGLYPSKSLISGVVLSDAKRFQINLCGGDIAFH 56
QY	268 LNPRFDENAVVRNTQIDNSGWSSEERSLPRKMPFVRGGSFWSWILCEAHLKVAVDGQHLF 327
DB	: : : : : : : : : : : : : : :
DB	57 LNPRFDENAVVRNTQINNSGWEPEERSLPGMSRGRGSFWSWILCEGHCFAVDGQHIC 116
QY	328 EYVHRLNRLPTINLRVGGDIQLTHVOT 355
DB	: : : : : : : : : : : : : : :
DB	117 EYSHRLMNLPTINLRVAGDIQLTHVET 144
RESULT 8	

ID LEG6_MOUSE STANDARD; PRT; 301 AA.
AC O54891; O88352;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GALECTIN-6.
GN LGALS6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112847; PubMed=9446608;
RA Gitt M.A., Colnot C., Poirier F., Nani K.J., Barondes S.H.,
RA Leffler H.;
RT "Galectin-4 and galectin-6 are two closely related lectins expressed
in mouse gastrointestinal tract.";
RL J. Biol. Chem. 273:2954-2960(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=98112848; PubMed=9446609;
RA Gitt M.A., Xia Y.-R., Atchison R.E., Lusis A.J., Barondes S.H.,
RA Leffler H.;
RT "Sequence, structure, and chromosomal mapping of the mouse Lgals6
gene, encoding galectin-6.";
RL J. Biol. Chem. 273:2961-2970(1998).
CC -!- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
DOMAINS.
CC -!- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AF026799; AAC04508.1;
DR EMBL; AF026796; AAC04508.1; JOINED.
DR EMBL; AF026797; AAC04508.1; JOINED.
DR EMBL; AF026798; AAC04508.1; JOINED.
DR EMBL; AF026794; AAC27244.1;
DR MGD; MGI:107535; Lgals6.
DR InterPro; IPR001079;
DR Pfam; PF00337; Gal-bind_lectin; 2.
DR Galaptin; Lectin; Repeat.
DR DOMAIN 1 151 GALAPTIN 1.
FT DOMAIN 152 160 LINKER.
FT DOMAIN 160 301 GALAPTIN 2.
FT CONFLICT 154 154 A -> V (IN AAC27244).
SQ SEQUENCE 301 AA; 34112 MW; 9A4DD09944EDEFAB9 CRC64;

Query Match 27.0%; Score 518; DB 1; Length 301;
Best Local Similarity 35.6%; Pred. No. 1.4e-34;
Matches 124; Conservative 52; Mismatches 110; Indels 62; Gaps 10;

QY 9 PVLSPAVPFGSTIQGLQDGLQITVNGTVLSSSGTRFAVNFOTG-FSGNDIAFHFNPRE 67
DB 11 PTYNPLPKRPITPGSLVGMSEFYIGTA-KENMRFRHNFVAVGDDGADVAFHFNPRE 69
QY 68 DGVVVCNTRQMSWGPEERKTHMPFQKMPFDLCFLVQSSDFKVMVNGILFVQYFHRYP 127
DB 70 GWDKVVFNKQSGRWCKEEKS-MPFQKQKHFLVPMVMEPEHYKVVVNGSPFVEYGHRLP 128
QY 128 HRVDTISVNGSVOLSYISFQNPRTVPVQPAFSTVPFSPVCPFPPRGRKPPGWFA 187
DB 129 VQMVTHLQVDGDLQSLINF-----FGVQPAETKYP----- 159

QY 188 NPAPITQTVIHTVQAPGQMFSTPAIPPMYPHPAYPMPFITITLGGLYPSKILLSGTV 247
DB 160 -----AMTG-----PPVFN-----CLPYVGALOGGFTVRTTIKGV 193
QY 248 LPSAORFHIN--LCSGNHIAFHLPFRDENAVVNTQIDNSWGSERSLPRKMPFVRGOS 305
DB 194 LPTAKTFAINFRVGSSEIDIALHINPRIGD-CLVRNSYMGNSWGTEEMVAYN-PFGPGQF 251
QY 306 FSVWILCEAHCLKVAVDGQHLFEYVHRLNLPINRLEVGVDIOLTHV 353
DB 252 FDLSTRCGMDRFKVFANGIHLEFNFSHRFOALRKINTLEINGDLTUSV 299
RESULT 9
LEG8_MOUSE STANDARD; PRT; 316 AA.
ID LEG8_MOUSE
AC Q9JUL15;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE GALECTIN-8 (LGALS-8).
GN LGALS8.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL6/J; TISSUE=Kidney;
RA Maier C., Haeussler J., Roesch K., Moschgath E., Haeussler J.,
RA Vogel W.;
RT "The human Lgals-8 gene: genomic sequence and expression of the
prostate carcinoma tumour antigen (PCTA-1) and the P066 carbohydrate
binding protein.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: POSSESSES SUGAR BINDING AND HEMAGGLUTINATION ACTIVITY.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; AF218069; AAF27645.1;
DR MGD; MGI:1928481; Lgals8.
DR InterPro; IPR001079;
DR PROSITE; PS00309; GALAPTIN; 1.
DR Galaptin; Lectin; Repeat.
DR DOMAIN 1 153 GALAPTIN 1.
FT DOMAIN 154 184 LINKER.
FT DOMAIN 185 316 GALAPTIN 2.
FT BINDING 248 254 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 316 AA; 36161 MW; 11A20309AEF52C69 CRC64;

Query Match 24.4%; Score 467.5; DB 1; Length 316;
Best Local Similarity 32.4%; Pred. No. 1.6e-30;
Matches 113; Conservative 61; Mismatches 124; Indels 51; Gaps 8;

QY 12 SPAPFSGTIQGLQDGLQITVNGTVLSSSGTRFAVNFOTGFS---GNDIAFHFNPRE 68
DB 13 NPILPYVGITITQLKPGSLIVIRGHVPKDS-ERFQDFQLGNSLKPADVAFHFNPRFR 71
QY 69 GGVVVCNTRQMSWGPEERKTHMPFQKMPFDLCFLVQSSDFKVMVNGILFVQYFHRVP 128
DB 72 SSCIVCNLTQKQWGEIEITYDMPFERKESFEIVFVNLKFKQVAVNGRHVLLYAHRI 131
QY 129 HRVDTISVNGSVOLSYISFQNPRTVPVQPAFSTVPFSPVCPFPPRGRKPPGWFA 188
DB 132 EQIDTVGIYKVINHSIGFR-----FSSDLQSM-----ETS 162

QY 189 PAPITQTVHTVQVAPGQMFSTPAIPPMYHPAYPMPFITITLGLGLYPSKSIILSGTVL 248
 DB 163 ALGLTQINRENIQK-PGKL-----QLSLPFEARLNASMGPGRTVVIKGEVN 207
 QY 249 PSAORFHINLCSG--NHIAFHLPNPFDENAVVNTQIDNSWGSSEERSLPRKMPFVRGQSF 306
 DB 208 TNARSFNVLDVAGTRDIALHLNPLNKAFRNSFLQDANGEBERNI-TCFPFSSGMYF 266
 QY 307 SWVLCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLTHVOT 355
 DB 267 EMIIYCDVREFKVAINGVHSLEYKHKRFDLSSIDTSLSDGDIRLLDVR 315

RESULT 10
 LEG8_RAT
 ID LEG8_HUMAN STANDARD; PRT: 316 AA.
 000214; 015215; Q9UP34; Q9UE26; Q9UP33; Q9UP32;
 01-NOV-1997 (Rel. 35, Created)
 01-OCT-2000 (Rel. 40, Last sequence update)
 01-OCT-2000 (Rel. 40, Last annotation update)
 DE GALECTIN-8 (GAL-8) (PROSTATE CARCINOMA TUMOR ANTIGEN) (PCTA-1)
 DE (PO66 CARBOHYDRATE-BINDING PROTEIN) (PO66-CBP).
 GN LGALS8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Prostate;
 RX MEDLINE=96293510; PubMed=8692978;
 RA Su Z.-Z., Lin J., Shen R., Fisher P.E., Goldstein N.I., Fisher P.B.;
 RT "Surface-epitope masking and expression cloning identifies the human
 RT prostate carcinoma tumor antigen gene PCTA-1 a member of the galectin
 RT gene family.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:7252-7257(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Hippocampus;
 RA Hadari Y.R., Eisenstein M., Zakut R., Zick Y.;
 RT "Galectin-8: on the road from structure to function.";
 RT Trends Glycosci. Glycotechnol. 9:103-112(1997).
 RN [3]
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RC TISSUE-Lung carcinoma;
 RA Brichory F., Biron N., Desrues B., Bourguet P., Le Pennec J.P.,
 RT Dazord L.;
 RT "Molecular cloning of a beta-galactoside-binding lectin related to
 RT galectin-8 and identified in human lung carcinoma.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Maier C., Haessler J., Roess K., Moschagath E., Vogel W.;
 RT "Genomic organization and expression of the human galectin-8 gene.";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
 CC -1- TISSUE SPECIFICITY: SELECTIVE EXPRESSION BY PROSTATE CARCINOMAS
 CC VERSUS NORMAL PROSTATE AND BENIGN PROSTATIC HYPERTROPHY.
 CC -1- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING
 CC DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L78132; AAB51605.1; .
 DR EMBL; X91790; CAA62904.1; ALT_INIT.

DR EMBL; AF074000; AAD45402.1; .
 DR EMBL; AF074001; AAD45403.1; .
 DR EMBL; AF074002; AAD45404.1; .
 DR EMBL; AF193806; AAF19370.1; ALT_INIT.
 DR EMBL; AF193805; AAF19370.1; JOINED.
 DR HSSP; P17931; IA3K.
 DR InterPro; IPR001079; .
 DR PROSITE; PS00337; Gal-bind_lectin; 2.
 DR GALAPTIN; Lectin; Repeat; Antigen; Alternative splicing; Polymorphism.
 FT DOMAIN 1 153 GALAPTIN 1.
 FT LINKER.
 FT DOMAIN 154 184 GALAPTIN 2.
 FT DOMAIN 185 316 BETA-LACTOSIDE (BY SIMILARITY).
 FT BINDING 248 254 L -> LPSNRGGDISKIAPRTVYTKSKDSTVNHFLTCTKIP
 FT VARSPLIC 182 182 PMNVSK (IN ISOFORM 2).
 FT VARSPLIC 182 182 L -> LQTVSPSWLDQGHGSETFCFVSLWTRVFLFIACRP
 FT VARSPLIC 182 182 GLTVASFO (IN ISOFORM 3).
 FT VARIANT 35 35 /FTID=VAR_009710.
 FT CONFLICT 18 18 Y -> F (IN REF. 1 AND 2).
 FT CONFLICT 55 55 V -> M (IN REF. 2).
 FT CONFLICT 97 99 KRE -> OKRE (IN REF. 2).
 FT CONFLICT 111 111 D -> A (IN REF. 2).
 FT CONFLICT 170 170 S -> V (IN REF. 1).
 FT CONFLICT 183 183 S -> R (IN REF. 2).
 FT CONFLICT 203 203 K -> Q (IN REF. 1).
 SQ SEQUENCE 316 AA; 35538 MW; 1724C36085A0510A CRC64;

Query Match 24.2%; Score 463.5; DB 1; Length 316;
 Best Local Similarity 31.8%; Pred. No. 3.3e-30;
 Matches 112; Conservative 52; Mismatches 131; Indels 57; Gaps 7;
 QY 12 SPAPVFGSTIOGGLODGLQITVNGTVLSSSTGTRFAVNPQTGFSGN---DIAFHENPRED 68
 DB 13 NPVIPPYGTIPDQLDPGLTIVICGHV-PSDADRFQDLQNGSSVKPRADVAFHENPREKR 71
 QY 69 GGYYVNCNTRONGSWGPERKTHMPGKMPDFLCFLQSSDFKVMVNGILFVQVHFHVPF 128
 DB 72 AGCIVCNTLINEKGREIYDTDFEKEKSEFIVIMVLKDKFQVAVNGKHTLLYGHRI 131
 QY 129 HRVDTISVNGSVQLSYISFQNPRTVPVQPAFTVPFSPQVCFPPRPRGRROKPGVWPAN 188
 DB 132 EKIDTLGIYGVNIHSIGF-----SPSS----- 154
 QY 189 PAPITQTVHTVQVAPGQMFSTPAIPPMYHP---PAYPMPFITITLGLGLYPSKSIILSG 245
 DB 155 -----DLQSTQASSLELTSIRENVKSGTPTQLSLPFAARLNTPMGRTVVVVK 204
 QY 246 TVLPFAORFHINLCSG--NHIAFHLPNPFDENAVVNTQIDNSWGSSEERSLPRKMPFVRG 303
 DB 205 EVNNAKSFNVLDLAGSKDIALHLNPLNKAFRNSFLQESWGEERNI-TSFPFSPG 263
 QY 304 QSFVSWILCEAHCLKVAVDGQHLFEYHRLNRLPTINRLEVGDDIQLTHVOT 355
 DB 264 MYFEMIYCDVREFKVAINGVHSLEYKHKRFDLSSIDTSLSDGDIRLLDVR 315

RESULT 11
 LEG8_RAT
 ID LEG8_RAT STANDARD; PRT: 316 AA.
 AC Q62665;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE GALECTIN-8 (30 KDA S-TYPE LECTIN) (RL-30).
 GN LGALS8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]

QY	132	DTISVNGSVOLSVISFQNTPTVPVQFAFSTVPFSQPVCPFPPRRGRKQPPGVPWNPANP	191
Db	128	SHLSIDGDLNLNHH	146
QY	192	ITQTVIHTVSQAFQMFSTPAIPPMYPHPAYPMPFTTTILGLGYPSKSTLLSGTVLPSA	251
Db	147	-----YYPVYESGLANGLPVQKSLVETGVEKKA	176
QY	252	QRFHINLCGN-HIAFHLNPRFDENAVVRNTQIDNSWGSEERSLPRKMPFVRQGSFVWI	310
Db	177	KREHVNLLRRNGDISFHFNPREFDEKHKVIRNSLAENWNEERE-GKNPPEKGVGFDLVI	234
QY	311	LCBAHCLKVAVDGQHLEFYHYHRLNPLPTINRLEVRGGDIQLTHVQ	354
Db	235	QNEEYAFQVFNVEGRIYSFAHR-ADPHDIAGLAGTSGDIELSGIQ	277

RESULT 14

ID	LE33_CAEEEL	STANDARD;	PRT;	285 AA.
AC	Q09581;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	32 KDA BETA-GALACTOSIDE-BINDING LECTIN LEC-3 (32 KDA GBP).			
GN	LEC-3 OR ZK892.1.			
OS	Caenorhabditis elegans.			
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;			
OC	Rhabditidae; Paloderinae; Caenorhabditis.			
OX	NCBI_TaxID=6239;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=BRISTOL N2;			
RC	STRAIN=BRISTOL N2;			
RA	Lloyd C.;			
RL	Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.			
RP	[2]			
RN	REVIEWS.			
RP	STRAIN=BRISTOL N2;			
RC	STRAIN=BRISTOL N2;			
RA	Jones S.J.M.;			
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.			
CC	- FUNCTION: BINDS GALACTOSE (BY SIMILARITY).			
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation the European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/annotation or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z48638; CAA88570.1; -			
DR	HSSP; P17931; 1A3K.			
DR	WormPep; ZK892.1; CE18468.			
DR	InterPro; IPR001079; -			
DR	Pfam; PF00337; Gal-bind_lectin; 2.			
DR	PROSITE; PS00309; GALAPTIN; 1.			
KW	Galaptin; Lectin; Repeat; Multigene family.			
FT	DOMAIN 1 137 GALAPTIN 1.			
FT	DOMAIN 138 285 GALAPTIN 2.			
FT	BINDING 212 218 BETA-GALACTOSIDE (BY SIMILARITY).			
FT	SEQUENCE 285 AA; 32465 MW; 5D191988B760B4AE CRC64;			

Query Match 19.5%; Score 374; DB 1; Length 285;
 Best Local Similarity 28.1%; Pred. No. 4.4e-23;
 Matches 97; Conservative 54; Mismatches 112; Indels 82; Gaps

QY	15	VPESGTIOGGLOGLOITVNCNTVLSSSGTFAVNF---QTGSGNDIAFHNPFEDGGY	71
Db	9	IPYRSKUTERIEPQTLLIRKGTIDES-KRFNINLHKDSDPFGNDVPLSHLIRF-DEGK	66
QY	72	VVCNTRONGSWGPERKTHMPFOKGMFEDLCFLVOSDFKVMYNGILFVOYFHRVPPHRV	131

Db 67 IVYNAYTKGTWKEERAKN-PIKKGDDDFDIRAHDSKFQVSNHKEVNFHRIPLNSV 125
QY 132 DTISVNGSVOLSVISQNPRTVPQAFSTVPFSQVCPFPGRGRQRQPPGWPNPAP 191
Db 126 SHLSIDGDDVVLNHVQ-----WGK-----144
QY 192 ITQTVHTVQSAPGOMFSTPAIPPMYPHPAYPMPFITILG-GLYPSKILLSGTVLPS 250
Db 145 -----YYPVYESGIAADGLVPGKTLVVYGTPEKK 174
QY 251 AQRFHNLCSGN-HIAFHLNPRFDENAVRNTQIDNSWSEERSLPRKMPFVGRGSFSVM 309
Db 175 AKKFENLLKKGKIDALHENPRFDEKSVSRNSLVNGEWNEERE--GKNPFERLTAFDLE 232
QY 310 ILCEAHCLVAVDGOHLFEYYHRLNLPINRLEVGGDIQLTHVQ 354
Db 233 IRNEFAFQIFVNGERFASYAHRV-DPHDIAGLIQGDIELTGIQ 276

RESULT 15

CRILLO STANDARD; PRT; 244 AA.
AC P47953;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE GALECTIN-3 (GALACTOSE-SPECIFIC LECTIN 3) (MAC-2 ANTIGEN) (IGE-BINDING
DE PROTEIN) (35 KDA LECTIN) (CARBOHYDRATE BINDING PROTEIN 35) (CBP 35)
DE (LAMININ-BINDING PROTEIN) (LECTIN L-29) (CBP30).
GN LGALS3.
OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetulus.
OX NCBI_TaxID=10030;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=94299546; PubMed=8027086;
RA Mehul B., Bawumia S., Martin S.R., Hughes R.C.;
RT "Structure of baby hamster kidney carbohydrate-binding protein CBP30,
RT an S-type animal lectin.";
RL J. Biol. Chem. 269:18250-18258(1994).
CC -!- FUNCTION: GALACTOSE-SPECIFIC LECTIN WHICH BINDS IGE.
CC -!- SIMILARITY: THE C-TERMINAL DOMAIN BELONGS TO THE GALAPTIN
CC (S-LECTIN) FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X78879; CAA55479.1; -
DR HSP: P17931; I43K.
DR InterPro: IPR001079; -
DR Pfam: PF00337; Gal-bind_lectin; 1.
DR PROSITE: PS00309; GALAPTIN; 1.
KW Galaptin; Lectin; Ige-binding protein; Repeat; Phosphorylation;
KW Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT MOD_RES 5 5 PHOSPHORYLATION (BY CK1)
FT FT (BY SIMILARITY).
FT DOMAIN 34 98 7 X 9 AA TANDEM REPEATS OF Y-P-G-X(3)-P-
FT [GS]-A.
FT REPEAT 34 42 1.
FT REPEAT 43 51 2.
FT REPEAT 52 60 3.
FT REPEAT 61 69 4.

FT REPEAT 70 77 5 (APPROXIMATE).
FT REPEAT 78 87 6 (APPROXIMATE).
FT REPEAT 88 98 7 (APPROXIMATE).
FT DOMAIN 112 244 GALAPTIN.
FT DISULFID 167 167 INTERCHAIN (BY SIMILARITY).
FT BINDING 175 181 BETA-GALACTOSIDE (BY SIMILARITY).
SQ SEQUENCE 244 AA; 25608 MW; 8F99B9AA0BBA7D3F CRC64;
Query Match 17.7%; Score 339; DB 1; Length 244;
Best Local Similarity 38.9%; Pred. No. 2.3e-20;
Matches 88; Conservative 26; Mismatches 84; Indels 28; Gaps 9;
QY 149 NRTVP-----VQPAESTVP-FSQPVCFPPR-PRGR--ROKPPGVWPA-----NPAPI 192
Db 17 NQGWPGAWGNQPGAGGYPCASYPGAYPCQAPPQAYPGQAPGAYPCPTAPGAYPCGAPG 76
QY 193 TOT----VIHTVOSAPGOMFSTPAIPPMYPHPAYPMPFITILGGLYPSKILLSGTVL 248
Db 77 AYPGQPGASGAYPSAPG---AYPAAGPYGAPTGAITVYKPLPLAGGVMPRLITINGTVK 133
QY 249 PSAQRFHNLCSGNHIAFHLNPRFDEN---AVVRNTQIDNSWSEERSLPRKMPFVRGQS 305
Db 134 PNANRIILNFLRGNDIAHFENPRFENNRRVIVCNTKQDNNGREERQ---SAPPFESGRP 191
QY 306 FSVWILCEAHCLKLVAVDGOHLFEYYHRLNLPINRLEVGGDIQLT 351
Db 192 FKIQVLVEADHFKVAVNDHALLIQYHNRKMLNREINOMEISGDITLT 237

Search completed: August 1, 2001, 09:42:19
Job time: 684 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 1, 2001, 09:30:55 ; Search time 36.57 Seconds
(without alignments)
29.975 Million cell updates/sec

Title: US-09-485-951-1
Perfect score: 178
Sequence: 1 NPRTVPVQPAFTVPFSPQVCPFRPRGRQRK 32

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Archived: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	178	100.0	355	1 LEG9_HUMAN	O00182 homo sapien
2	104	58.4	354	1 LEG9_RAT	P97840 rattus norv
3	94	52.8	353	1 LEG9_MOUSE	O08573 mus musculu
4	62.5	35.1	806	1 MK07_MOUSE	Q9wvs8 mus musculu
5	55	30.9	1490	1 CRK7_HUMAN	Q9yvv4 homo sapien
6	54	30.3	487	1 EBN2_EBV	P12978 epstein-bar
7	53.5	30.1	360	1 N044_SOYBN	P04672 glycine max
8	52.5	29.5	791	1 ARBT_MOUSE	P53762 mus musculu
9	51.5	28.9	169	1 HSB7_MOUSE	P53385 mus musculu
10	51.5	28.9	970	1 PSU1_YEAST	P53550 saccharomyc
11	51	28.7	460	1 MYCN_MARMO	Q61976 marmota mon
12	51	28.7	1172	1 TSP2_HUMAN	P35442 homo sapien
13	51	28.7	1429	1 L112_CABEL	P14585 caenorhabdi
14	50.5	28.4	90	1 HSB7_RAT	Q9uk5 rattus norv
15	50.5	28.4	489	1 CPCM_RAT	P19225 rattus norv
16	50.5	28.4	800	1 ARNT_RAT	P41739 rattus norv
17	50	28.1	1139	1 RBL2_HUMAN	O08999 homo sapien
18	50	28.1	1400	1 RON_HUMAN	Q04912 homo sapien
19	49.5	27.8	324	1 GC1_MOUSE	P01868 mus musculu
20	49.5	27.8	393	1 GC1M_MOUSE	P01869 mus musculu
21	49.5	27.8	464	1 MYCN_HUMAN	P04198 homo sapien
22	49.5	27.8	551	1 IL2B_HUMAN	P14784 homo sapien
23	49.5	27.8	872	1 S145_HUMAN	Q13435 homo sapien
24	49	27.5	417	1 HEM2_SELMA	P45623 selaginella
25	49	27.5	434	1 CPCO_RAT	P33273 rattus norv
26	49	27.5	1135	1 RBL2_MOUSE	Q64700 mus musculu
27	49	27.5	1135	1 RBL2_RAT	O5081 rattus norv
28	49	27.5	1175	1 DSRA_RAT	P55266 rattus norv
29	49	27.5	1906	1 KMLS_CHICK	P11799 gallus gall
30	48.5	27.2	170	1 HSB7_HUMAN	Q9ub99 homo sapien
31	48.5	27.2	324	1 MATD_NEOCR	Q10116 neurospora
32	48.5	27.2	524	1 FTWH_MYCTU	O06223 mycobacteri
33	48.5	27.2	790	1 ARNT_RABIT	O02748 oryctolagus

RESULT 1

LEG9_HUMAN					
ID	LEG9_HUMAN	STANDARD;	PRT;	355 AA.	
AC	O00182; O14532; O75028;				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	01-OCT-2000 (Rel. 40, Last annotation update)				
DE	GALECTIN-9 (HOM-HD-21) (ECALECTIN).				
GN	LGALS9.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A. (SHORT FORM).				
RC	TISSUE=Splice;				
RX	MEDLINE=97197815; PubMed=9045665;				
RA	Tuerceci O., Schmitt H., Fadle N., Pfreundschuh M., Sahin U.;				
RT	"Molecular definition of a novel human galectin which is immunogenic				
RT	in patients with Hodgkin's disease.";				
RL	J. Biol. Chem. 273:16976-16984(1998).				
RN	[2]				
RP	SEQUENCE FROM N.A. (LONG FORM).				
RC	TISSUE=Gastric carcinoma;				
RA	Kato S.;				
RL	Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A. (SHORT FORM).				
RX	MEDLINE=98307937; PubMed=9642261;				
RA	Matsumoto R., Matsumoto H., Seki M., Asano Y., Kanegasaki S.,				
RA	Stevens R.L., Hirashima M.;				
RT	"Human ecalectin, a variant of human galectin-9, is a novel eosinophil				
RT	chemoattractant produced by T lymphocytes.";				
RL	J. Biol. Chem. 273:16976-16984(1998).				
CC	- FUNCTION: BINDS GALACTOSIDES. MAY PLAY A ROLE IN THYMOCYTE-				
CC	EPITHELIAL INTERACTIONS RELEVANT TO THE BIOLOGY OF THE THYMUS. THE				
CC	SHORT ISOFORM ACTS AS AN EOSINOPHIL CHEMOATTRACTANT.				
CC	- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS; A LONG FORM (SHOWN				
CC	HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.				
CC	- TISSUE SPECIFICITY: PERIPHERAL BLOOD LEUKOCYTES AND LYMPHATIC				
CC	TISSUES. OVEREXPRESSED IN HODGKIN'S DISEASE TISSUE.				
CC	- DOMAIN: CONTAINS TWO HOMOLOGOUS BUT DISTINCT CARBOHYDRATE-BINDING				
CC	DOMAINS.				
CC	- SIMILARITY: BELONGS TO THE GALAPTIN (S-LECTIN) FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
CC	EMBL; 249107; CAA88922.1;				
DR	EMBL; AB006782; BAA22166.1;				
DR	EMBL; AB005894; BAA31542.1;				

Q10370 glycine max
P09513 barley yell
P04727 triticum ae
P54704 dictyosteli
O19012 erinaceus e
P08328 escherichia
P18089 homo sapien
Q15654 homo sapien
Q05649 mus musculu
P54734 anabaena sp
O94751 schizosacch
Q00423 glycine max

ALIGNMENTS

CC -1- PTM: AUTOPHOSPHORYLATED ON THREONINE AND TYROSINE RESIDUES, WHEN
CC THE C-TERMINAL PART OF THE KINASE, WHICH COULD HAVE A REGULATORY
CC ROLE, IS ABSENT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC MAP KINASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB019373; BAA82039.1; -
CC HSSP; P27703; 2ERK.
CC MGD; MGI:1346347; MapK7.
CC InterPro; IPR000719; -
CC InterPro; IPR002290; -
CC DR. PROSITE; PS01351; MAPK; 1.
CC DR. PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC DR. PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
CC DR. PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC KW Transferase; Serine/threonine-protein kinase; ATP-binding; Cell cycle;
CC KW Phosphorylation.
CC FT DOMAIN 55 347 PROTEIN KINASE.
CC FT DOMAIN 434 465 PRO-RICH 1.
CC FT DOMAIN 521 524 POLY-ARG.
CC FT DOMAIN 578 700 PRO-RICH 2.
CC FT NP_BIND 61 69 ATP (BY SIMILARITY).
CC FT BINDING 84 84 ATP (BY SIMILARITY).
CC FT ACT_SITE 182 182 BY SIMILARITY.
CC FT MOD_RES 219 219 PHOSPHORYLATION (ACTIVATES THE KINASE)
CC FT (BY SIMILARITY).
CC FT MOD_RES 221 221 PHOSPHORYLATION (ACTIVATES THE KINASE)
CC FT (BY SIMILARITY).
CC SQ SEQUENCE 806 AA; 87732 MW; E7CC41C4BBD0633 CRC64;

Query Match 35.1%; Score 62.5; DB 1; Length 806;
Best Local Similarity 62.5%; Pred.No. 1.6;
Matches 15; Conservative 1; Mismatches 7; Indels 1; Gaps 1;

QY 4 TVPVQPAFTV-PFSQPVCFPPRP 26
I : I I I I I I I I I I I I I I I I
Db 610 TGPLQPAISIPGPASQPVCPVPPGP 633

RESULT 5
CRK7_HUMAN STANDARD; PRT; 1490 AA.
ID ID CRK7_HUMAN
AC Q9NYV4; O94978;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CELL DIVISION CYCLE 2-RELATED PROTEIN KINASE 7 (EC 2.7.1.-) (CDC2-
DE RELATED PROTEIN KINASE 7) (CKRKS).
DE CRK7 OR KIAA0904.
GN GN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pines J.N., Kelly E.;
RT "CRK7: a novel CDC2-related protein kinase that colocalizes with
RL interchromatin granule clusters.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 266-1262 FROM N.A.
RP
RC TISSUE=Brain;
RX MIMLN1=99156230; PubMed=10048485;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirotsawa M.

This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

EMBL: U10325; AAA56717.1; -;
 EMBL: U14333; AAA61732.1; -;
 HSP: P22415; 1AN4.
 MGD: MGI:88071; Aint.
 InterPro: IPR000014; -;
 InterPro: IPR001067; -;
 InterPro: IPR001092; -;
 InterPro: IPR003015; -;
 Pfam: PF00010; HLH; 1;
 Pfam: PF00989; PAS; 2;
 PRINTS: PR00785; NCTRNSLOCATR.
 PROSITE: PS00038; HELIX_LOOP_HELIX; 1;
 Nuclear protein; DNA-binding; Transcription regulation; Activator;
 Alternative splicing; Repeat.

FT DNA_BIND 90 102 BASIC DOMAIN
 FT DOMAIN 103 143 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT REPEAT 163 230 PAS-1.
 FT REPEAT 351 417 PAS-2.
 FT DOMAIN 424 467 PAC MOTIF.
 FT DOMAIN 99 102 POLY-ARG.
 FT DOMAIN 503 507 POLY-GLN.
 FT DOMAIN 740 743 POLY-SER.
 FT VARSPIC 77 100 SDDEQSSADSKERLARENHSEIERR -> TKFL (IN
 FT SHORT ISOFORM).
 FT CONFLICT 411 411 S -> T (IN REF. 2).
 FT CONFLICT 534 534 A -> R (IN REF. 2).
 FT CONFLICT 644 644 T -> S (IN REF. 2).
 FT CONFLICT 650 650 A -> C (IN REF. 2).
 FT SEQUENCE 791 AA; 86977 MW; BDEB79E4BD75D452 CRC64;

Query Match 29.5%; Score 52.5; DB 1; Length 791;
 Best Local Similarity 35.7%; Pred. No. 27;
 Matches 15; Conservative 3; Mismatches 9; Indels 15; Gaps 2;

QY 1 NRPRTVPQPAF-----STVP-----FSQVPCFPFPRR 27
 DB 556 DFRPELPYSITADQSGISSIVPATQQLFSQSSSFPNPR 597
 :||:|:||||| ||| ||| ||| |||
 :||:|:||||| ||| ||| ||| |||

RESULT 9
 HSB7_MOUSE STANDARD; PRT; 169 AA.
 ID HSB7_MOUSE
 AC P35385; Q9QUS2;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HEAT-SHOCK PROTEIN, BETA-7 (CARDIOVASCULAR HEAT SHOCK PROTEIN) (CVHSP)
 DE (HEAT SHOCK PROTEIN 25 KDA 2) (PROTEIN P19/6.8).
 DE HSPB7 OR HSP25-2 OR CVHSP.
 GN Mus musculus (Mouse).
 OS
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=20062883; PubMed=10593960;
 RA Krief S., Faivre J.-F., Robert P., Le Douarin B., Brument-Larignon N.,
 RA Lefrere I., Bouzyk M.M., Anderson K.M., Grellier L.D., Tobin F.L.,
 RA Souchet M., Brill A.;
 RT Identification and characterization of cvhsp. A novel human small
 RT stress protein selectively expressed in cardiovascular and
 RT insulin-sensitive tissues.;
 RL J. Biol. Chem. 274:36592-36600(1999).

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X53673; CAA37712.1; -
 CC EMBL; X53674; CAA37712.1; JOINED.
 DR InterPro; IPR001092; -
 DR InterPro; IPR002418; -
 DR InterPro; IPR003015; -
 DR Pfam; PF00010; HLH; 1.
 DR Pfam; PF01056; Myc_N_term; 1.
 DR PRINTS; PR00044; LEUZIPPRMYC.
 DR PROSITE; PS00038; HELIX_LOOP_HELIX; 1.
 KW Nuclear protein; DNA-binding; Proto-oncogene; Phosphorylation.
 FT DOMAIN 258 274 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 377 390 BASIC DOMAIN.
 FT DOMAIN 391 430 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
 FT DOMAIN 429 450 LEUCINE-ZIPPER (POTENTIAL).
 FT MOD_RES 257 257 PHOSPHORYLATION (BY CK2)
 (BY SIMILARITY).
 FT MOD_RES 259 259 PHOSPHORYLATION (BY CK2)
 (BY SIMILARITY).
 FT SEQUENCE 460 AA; 49192 MW; 8A16686C82F5B02E CRC64;

Query Match 28.7%; Score 51; DB 1; Length 460;
 Best Local Similarity 38.5%; Pred. No. 24;
 Matches 15; Conservative 1; Mismatches 13; Indels 10; Gaps 1;

QY 1 NPTVPVQPAFS-----TVPFSQVCPFPFRGR 29
 : | | | | | | | | | | | | | | | | | | | |
 Db 200 DPAPVPVAPGSPAVGAAGAAAPASAAVAPPLRGR 238

RESULT 12
 TSP2_HUMAN
 ID TSP2_HUMAN STANDARD; PRT; 1172 AA.
 AC P35442;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE THROMBOSPONDIN 2 PRECURSOR.
 GN THBS2 OR TSP2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 AC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94010892; PubMed=8406456;
 RA Labell T.L., Byers P.H.;
 RT "Sequence and characterization of the complete human thrombospondin 2
 cDNA: potential regulatory role for the 3' untranslated region.";
 RL Genomics 17:225-229(1993).
 RN [2]
 RP SEQUENCE OF 560-1172 FROM N.A.
 RC TISSUE=Fibroblast;
 RX MEDLINE=92217961; PubMed=1559694;
 RA Labell T.L., McGookey Milewicz D.J., Distèche C.M., Byers P.H.;
 RT "Thrombospondin II: partial cDNA sequence, chromosome location, and
 expression of a second member of the thrombospondin gene family in
 humans.";
 RL Genomics 12:421-429(1992).
 CC -!- FUNCTION: ADHESIVE GLYCOPROTEIN THAT MEDIATES CELL-TO-CELL AND
 CELL-TO-MATRIX INTERACTIONS. CAN BIND TO FIBRINOGEN, FIBRONECTIN,
 LAMININ AND TYPE V COLLAGEN.
 CC -!- SUBUNIT: HOMOTRIMER. CROSS-LINKED BY DISULFIDE BONDS.
 CC -!- SIMILARITY: BELONGS TO THE THROMBOSPONDIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 TSP TYPE-1 DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TSP TYPE-3 DOMAINS.

CC ----- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; L12350; AAA03703.1; -
 DR EMBL; M81339; -; NOT_ANNOTATED_CDS.
 DR PIR; A42173; A42173.
 DR HSP; P00740; LIIXA.
 DR MIM; 188061; -
 DR InterPro; IPR000561; -
 DR InterPro; IPR000884; -
 DR InterPro; IPR001007; -
 DR Pfam; PF00008; EGF; 2.
 DR Pfam; PF00090; tsp_1; 3.
 DR Pfam; PF00093; vwc; 1.
 DR PROSITE; PS00022; EGF_1; FALSE_NEG.
 DR PROSITE; PS01186; EGF_2; 1.
 DR PROSITE; PS50092; TSP1; 3.
 DR PROSITE; PS01208; VWFC; 1.
 KW Glycoprotein; Cell adhesion; Calcium-binding; Heparin-binding; Repeat;
 KW EGF-like domain; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 1172
 FT DOMAIN 19 232 THROMBOSPONDIN 2.
 FT DOMAIN 318 375 HEPARIN-BINDING (POTENTIAL).
 FT DOMAIN 381 432 VWFC.
 FT DOMAIN 437 493 TSP TYPE-1 1.
 FT DOMAIN 494 548 TSP TYPE-1 2.
 FT DOMAIN 549 589 TSP TYPE-1 3.
 FT DOMAIN 590 647 EGF-LIKE 1.
 FT DOMAIN 648 692 EGF-LIKE 2.
 FT DOMAIN 725 760 CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 761 783 EGF-LIKE 3.
 FT DOMAIN 784 819 TSP TYPE-3 1.
 FT DOMAIN 820 842 TSP TYPE-3 2.
 FT DOMAIN 843 880 TSP TYPE-3 3.
 FT DOMAIN 881 916 TSP TYPE-3 4.
 FT DOMAIN 917 952 TSP TYPE-3 5.
 FT DOMAIN 953 1172 TSP TYPE-3 6.
 FT SITE 928 930 C-TERMINAL.
 FT SITE 928 930 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 266 266 INTERCHAIN (PROBABLE).
 FT DISULFID 270 270 INTERCHAIN (PROBABLE).
 FT DISULFID 553 564 BY SIMILARITY.
 FT DISULFID 558 574 BY SIMILARITY.
 FT DISULFID 577 588 BY SIMILARITY.
 FT DISULFID 594 610 BY SIMILARITY.
 FT DISULFID 601 619 BY SIMILARITY.
 FT DISULFID 622 646 BY SIMILARITY.
 FT DISULFID 652 665 BY SIMILARITY.
 FT DISULFID 659 678 BY SIMILARITY.
 FT DISULFID 680 691 BY SIMILARITY.
 FT CARBOHYD 151 151 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 316 316 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 330 330 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 584 584 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 710 710 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 1069 1069 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 1172 AA; 129955 MW; 2AC7BB230E44C6F5 CRC64;

Query Match 28.7%; Score 51; DB 1; Length 1172;
 Best Local Similarity 44.8%; Pred. No. 60;
 Matches 13; Conservative 2; Mismatches 8; Indels 6; Gaps 2;

QY 2 PRTVPVQPAFSTVPFSQVCPFPFRGR 30
 : | | | | | | | | | | | | | | | | | | | |
 Db 608 PRCVNTQPGFHCLP-----C-PPRYRGNG 630

RESULT 13
 L112_CABEL
 ID L112_CABEL STANDARD; PRT; 1429 AA.
 AC P14595;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DE 01-OCT-2000 (Rel. 40, Last annotation update)
 DE LIN-12 PROTEIN PRECURSOR.
 GN LIN-12 OR R107.8
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RC STRAIN-BRISTOL N2;
 RX MEDLINE=88334747; PubMed=3419531;
 RA Yochem J., Weston K., Greenwald I.;
 RT "The Caenorhabditis elegans lin-12 gene encodes a transmembrane
 protein with overall similarity to Drosophila Notch.";
 Nature 335:547-550(1988).
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RC MEDLINE=94150718; PubMed=7906398;
 RA Willson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins J., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
 RA Sims M., Smaldon N., Smith A., Smith M., Sonnenhammer E., Staden K.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans";
 Nature 368:32-38(1994).
 RL Nature 368:32-38(1994).
 CC -1- FUNCTION: LIN-12 IS INVOLVED IN SEVERAL CELL FATES DECISIONS THAT
 CC REQUIRES CELL-CELL INTERACTIONS. IT IS POSSIBLE THAT LIN-12
 CC ENCODES A MEMBRANE-BOUND RECEPTOR FOR A SIGNAL THAT ENABLES
 CC EXPRESSION OF THE VENTRAL UTERINE PRECURSOR CELL FATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- SIMILARITY: HIGH. TO C.ELEGANS GLP-1.
 CC -1- SIMILARITY: CONTAINS 13 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
 CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 CC EMBL; M12069; AAA70191.1; -;
 CC EMBL; Z14092; CAA78474.1; -;
 CC PIR; S06434; S06434.
 CC HSSP; P00740; L1XA.
 CC WormPep; R107.8; CE00274.
 CC InterPro; IPR000152; -;
 CC InterPro; IPR000561; -;
 CC InterPro; IPR000800; -;
 CC InterPro; IPR001881; -;
 CC InterPro; IPR002110; -;
 CC Pfam; PF00008; EGF; 13.
 CC Pfam; PF00023; ank; 4.
 CC Pfam; PF00066; notch; 3.
 CC PROSITE; PS50088; ANK_REPEAT; 3.

 DR PROSITE; PS50297; ANK_REPEAT; 1.
 DR PROSITE; PS0010; ASX_HYDROXYL; 3.
 DR PROSITE; PS0022; EGF_1; 12.
 DR PROSITE; PS0186; EGF_2; 11.
 DR PROSITE; PS0187; EGF_CA; 2.
 KW Differentiation; Repeat; ANK repeat; EGF-like domain; Transmembrane;
 KW Glycoprotein; Signal.
 FT SIGNAL 1 15
 FT CHAIN 16 1429
 FT LIN-12 PROTEIN.
 FT DOMAIN 16 908
 FT EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 909 931
 FT POTENTIAL.
 FT DOMAIN 932 1429
 FT CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 20 61
 FT EGF-LIKE 1.
 FT DOMAIN 114 150
 FT EGF-LIKE 2.
 FT DOMAIN 152 190
 FT EGF-LIKE 3.
 FT DOMAIN 201 246
 FT EGF-LIKE 4.
 FT DOMAIN 250 285
 FT EGF-LIKE 5.
 FT DOMAIN 287 323
 FT EGF-LIKE 6.
 FT DOMAIN 323 363
 FT EGF-LIKE 7.
 FT DOMAIN 365 402
 FT EGF-LIKE 8.
 FT DOMAIN 404 441
 FT EGF-LIKE 9.
 FT DOMAIN 449 492
 FT EGF-LIKE 10.
 FT DOMAIN 503 541
 FT EGF-LIKE 11.
 FT DOMAIN 543 579
 FT EGF-LIKE 12.
 FT DOMAIN 582 619
 FT EGF-LIKE 13.
 FT REPEAT 635 669
 FT LIN/NOTCH 1.
 FT REPEAT 670 710
 FT LIN/NOTCH 2.
 FT REPEAT 711 750
 FT LIN/NOTCH 3.
 FT REPEAT 1093 1122
 FT ANK 1.
 FT REPEAT 1126 1158
 FT ANK 2.
 FT REPEAT 1162 1194
 FT ANK 3.
 FT REPEAT 1206 1236
 FT ANK 4.
 FT REPEAT 1240 1269
 FT ANK 5.
 FT DISULFID 24 35
 FT BY SIMILARITY.
 FT DISULFID 29 49
 FT BY SIMILARITY.
 FT DISULFID 51 60
 FT BY SIMILARITY.
 FT DISULFID 118 129
 FT BY SIMILARITY.
 FT DISULFID 123 138
 FT BY SIMILARITY.
 FT DISULFID 140 149
 FT BY SIMILARITY.
 FT DISULFID 156 169
 FT BY SIMILARITY.
 FT DISULFID 163 178
 FT BY SIMILARITY.
 FT DISULFID 180 189
 FT BY SIMILARITY.
 FT DISULFID 205 227
 FT BY SIMILARITY.
 FT DISULFID 221 234
 FT BY SIMILARITY.
 FT DISULFID 236 245
 FT BY SIMILARITY.
 FT DISULFID 254 264
 FT BY SIMILARITY.
 FT DISULFID 259 273
 FT BY SIMILARITY.
 FT DISULFID 275 284
 FT BY SIMILARITY.
 FT DISULFID 291 302
 FT BY SIMILARITY.
 FT DISULFID 296 311
 FT BY SIMILARITY.
 FT DISULFID 313 322
 FT BY SIMILARITY.
 FT DISULFID 327 339
 FT BY SIMILARITY.
 FT DISULFID 334 351
 FT BY SIMILARITY.
 FT DISULFID 353 362
 FT BY SIMILARITY.
 FT DISULFID 369 381
 FT BY SIMILARITY.
 FT DISULFID 375 390
 FT BY SIMILARITY.
 FT DISULFID 392 401
 FT BY SIMILARITY.
 FT DISULFID 408 419
 FT BY SIMILARITY.
 FT DISULFID 413 429
 FT BY SIMILARITY.
 FT DISULFID 431 440
 FT BY SIMILARITY.
 FT DISULFID 507 518
 FT BY SIMILARITY.
 FT DISULFID 512 529
 FT BY SIMILARITY.
 FT DISULFID 531 540
 FT BY SIMILARITY.
 FT DISULFID 547 558
 FT BY SIMILARITY.
 FT DISULFID 552 567
 FT BY SIMILARITY.
 FT DISULFID 569 578
 FT BY SIMILARITY.
 FT DISULFID 586 597
 FT BY SIMILARITY.
 FT DISULFID 591 607
 FT BY SIMILARITY.
 FT DISULFID 609 618
 FT BY SIMILARITY.
 FT CARBOHYD 41 41
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 165 165
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 194 194
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 378
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 515 515 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 751 751 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 754 754 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1429 AA; 157115 MW; 255EDD7A62C025DB CRC64;

Query Match 28.7%; Score 51; DB 1; Length 1429;

Best Local Similarity 54.5%; Pred. No. 72;

Matches 12; Conservative 0; Mismatches 6; Indels 4; Gaps 1;

QY 12 STVPFSQPVCF-----PPRPRGR 29

DB 196 STVEFKQPVCFLEISADHPDGR 217

RESULT 14

ID HSB7_RAT STANDARD; PRT; 90 AA.
 AC Q90UK5;
 DT 01-OCT-2000 (Rel. 40, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE HEAT-SHOCK PROTEIN, BETA-7 (CARDIOVASCULAR HEAT SHOCK PROTEIN) (CVHSP)
 DE (FRAGMENT).
 GN HSPB7 OR CVHSP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=20062883; PubMed=10593960;
 RA Krief S., Falvire J.-P., Robert P., Le Douarin B., Brument-Larignon N.,
 RA Lefrere I., Bouzyk M.M., Anderson K.M., Greller L.D., Tobin F.L.,
 RA Souchet M., Brill A.;
 RT Identification and characterization of cvhsp. A novel human small
 RT stress protein selectively expressed in cardiovascular and
 RT insulin-sensitive tissues.;
 RL J. Biol. Chem. 274:36592-36600(1999).
 CC -!- SUBUNIT: INTERACTS WITH C-TERMINAL DOMAIN OF ACTIN-BINDING PROTEIN
 CC 280 (BY SIMILARITY).
 CC -!- TISSUE SPECIFICITY: FOUND IN BOTH CARDIAC AND SKELETAL MUSCLE.
 CC -!- SIMILARITY: BELONGS TO THE SMALL HEAT SHOCK PROTEIN (HSP20)
 CC FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 DR EMBL; AJ243193; CAB63268.1; -;
 DR EMBL; AF155910; AAF20024.1; -;
 DR PROSITE; PS01031; HSP20; PARTIAL.
 KW Heat shock; Chapterone.
 FT NON_TER 1 8 POLY-SER.
 FT DOMAIN 1 8
 FT NON_TER 90 90
 SQ SEQUENCE 90 AA; 9804 MW; 9D122DB6443ED1FE CRC64;

Query Match 28.4%; Score 50.5; DB 1; Length 90;

Best Local Similarity 33.3%; Pred. No. 5.7;

Matches 11; Conservative 6; Mismatches 7; Indels 9; Gaps 1;

QY 6 PVQPAFS-----TVPFSPQVCFPPRGR 29

DB 17 PMEKALSMFSEDFGSMPLHSEPLTFPPRPGQ 49

RESULT 15

ID CPCM_RAT STANDARD; PRT; 489 AA.
 AC P19225;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE CYTOCHROME P450 2C22 (EC 1.14.14.1) (CYP11C22) (P450 MD) (P450 P49).
 GN CYP2C22 OR CYP2C-22 OR P450MD.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=90370497; PubMed=2395662;
 RA Nagata K., Sasamura H., Miyata M., Shimada M., Yamazoe Y., Kato R.;
 RT "cDNA and deduced amino acid sequences of a male dominant P-450md
 RT mRNA in rats";
 RL Nucleic Acids Res. 18:4934-4934(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Liver;
 RX MEDLINE=91088589; PubMed=2263625;
 RA Emi Y., Chijiwa C., Omura T.;
 RT "A different cytochrome P450 form is induced in primary cultures of
 RT rat hepatocytes";
 RL Proc. Natl. Acad. Sci. U.S.A. 87:9746-9750(1990).
 CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. IN LIVER MICROSOMES, THIS ENZYME IS INVOLVED IN AN
 CC NADPH-DEPENDENT ELECTRON TRANSPORT PATHWAY. IT OXIDIZES A VARIETY
 CC OF STRUCTURALLY UNRELATED COMPOUNDS, INCLUDING STEROIDS, FATTY
 CC ACIDS, AND XENOBIOTICS.
 CC -!- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
 CC OXIDIZED FLAVOPROTEIN + H(2)O.
 CC -!- SUBCELLULAR LOCATION: MEMBRANE-BOUND. ENDOPLASMIC RETICULUM.
 CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).

 DR EMBL; X53477; CAA37570.1; -;
 DR EMBL; M58041; AAA40950.1; -;
 DR PIR; S11160; S11160.
 DR PIR; A39257; A39257.
 DR InterPro; IPR001128; -;
 DR InterPro; IPR002401; -;
 DR Pfam; PF00067; p450; 1.
 DR PRINTS; PR00385; P450.
 DR PRINTS; PR00463; EP4501.
 DR PROSITE; PS00086; CYTOCHROME_P450; 1.
 KW Oxidoreductase; Monooxygenase; Electron transport; Membrane; Heme;
 KW Microsome; Endoplasmic reticulum.
 FT BINDING 434 434 HEME.
 FT CONFLICT 50 50 I -> M (IN REF. 2).
 FT CONFLICT 177 177 V -> L (IN REF. 2).
 FT CONFLICT 456 456 T -> I (IN REF. 2).
 FT CONFLICT 478 478 V -> L (IN REF. 2).
 FT CONFLICT 483 483 E -> Q (IN REF. 2).
 SQ SEQUENCE 489 AA; 56156 MW; E4C5B082AB81CB4D CRC64;

Query Match

Best Local Similarity 28.4%; Score 50.5; DB 1; Length 489;

Matches 12; Conservative 2; Mismatches 10; Indels 3; Gaps 1;

Qy 1 NPR--TVPQPAFSTVPFQVCFPP 24
| | | | | : | | |
Db 462 NPEDIDTTPVQPGLLSVPPPELGFIP 488

Search completed: August 1, 2001, 09:42:18
Job time: 683 sec

OM of: US-09-485-951-1 to: EST:* out_format : pfs

Date: Aug 1, 2001 12:16 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL-frame+p2n.model -DEV-xlp
-Q/cg2_1/USPTO.spool/US09485951/runat_31072001_202942_13211/app_query.fasta_1.504
-DB=EST -QMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEXT=0.000 -LOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -GAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-LIST=7.000 -START=1 -MATRIX=blom62 -TRANS=human40.cdi
-ALIGN=15 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
MAXLEN=2000000000 -USER=US09485951_GCGN1_1_3162 -NCPU=6
ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-485-951-1
Query length: 32
Database: EST:*
Database sequences: 10228115
Database length: 431459454
Search time (sec): 1962.150000

score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_est46:AW403229	+	178.00	370.43	1.6e-11	439	AW403229 UI-HF-BKO-abb-a-03-0-0
gb_est53:AW86539	+	171.00	350.83	2.0e-10	780	AW86539 RC1-OP0083-220300-021-
gb_gss9:AO279515	-	170.00	354.59	1.2e-10	394	AQ279515 C17B1-EL-25111111.TR CI
gb_est46:AW373817	+	164.00	337.52	1.1e-09	665	AW373817 QV3-BF0537-221299-048-
gb_est29:AL559254	+	152.00	314.23	2.2e-08	535	AL559254 AL559254 LTI-NFL008.TC
gb_est39:BF883419	+	118.00	245.36	0.0002	404	BF883419 QV3-ET0203-081200-525-
gb_est84:BF143767	+	106.00	215.45	0.0070	703	BF143767 601789654F1 NCI-CGAP-I
gb_est84:BF181157	+	106.00	213.25	0.0132	909	BF181157 601805820F1 NCI-CGAP-M
gb_est89:BF533977	+	106.00	211.12	0.0122	1165	BF533977 602075139F1 NCI-CGAP-
gb_est82:BF046316	+	104.00	217.93	0.0051	323	BF046316-BP250002B10C9 Soares
gb_est50:AW55379	+	104.00	213.54	0.0089	539	AW55379 106111 MARC IBOV Bos t
gb_est8:AA521753	+	85.00	172.93	1.63	590	AA521753 v115h07.1 R barstead m
gb_est7:BB144606	+	79.00	165.41	4.28	327	BB144606 BB144606 RIKEN full-1e
gb_est83:BF127246	+	76.50	151.11	26.81	941	BF127246 601650907R1 NIH-MGC-7
gb_est100:BG402468	+	70.50	137.34	156.72	1081	BG402468 60246617F1 NIH-MGC-7
gb_est57:BB133013	+	69.50	148.51	37.40	230	BB133013 BB133013 RIKEN full-1e
gb_gss7:AO123031	-	69.00	141.27	94.75	474	AQ123031 HS_3095-B1-B05-MF CIT
gb_est72:BE232585	+	68.50	146.20	50.34	236	BE232585 137643 MARC LPIC Sus s
gb_est67:BB516797	+	68.50	144.20	65.06	298	BB516797 BB516797 RIKEN full-1e
gb_est21:AT500463	-	68.50	139.93	112.41	490	AT500463 tn97e01.x1 NCI-CGAP-UT
gb_est86:BE307423	-	68.50	134.40	228.49	934	BF307423 601894071F1 NIH-MGC-17
gb_est90:BF588692	+	68.00	142.75	78.28	312	BF588692 7152e07.x1 NCI-CGAP-BI
gb_est81:BE980038	-	68.00	135.78	191.53	704	BE980038 UI-M-CG0-bcr-e-10-0-UI
gb_est73:BE359717	+	67.50	140.34	106.72	366	BE359717 DGI_56_E11.92.A002 Dar
gb_est90:BF600042	-	66.50	135.92	188.15	480	BF600042 264492 MARC 3BOV Bos t
gb_est80:BE866656	+	66.50	131.72	322.25	783	BE866656 60179255F1 NIH-MGC-53
gb_est85:BF217720	-	66.50	126.92	596.61	1371	BF217720 601822545F1 NIH-MGC-5
gb_est82:BF029198	-	66.00	129.96	403.96	851	BF029198 601765414F1 NIH-MGC-53
gb_est80:BE899224	-	66.00	128.62	479.73	995	BE899224 601682045F1 NIH-MGC-9
gb_est38:AV434080	+	65.00	132.83	279.68	477	AV434080 AV434080 Porphyra yezc
gb_est38:AV434616	+	65.00	132.42	294.54	500	AV434616 AV434616 Porphyra yezc
gb_est38:AV432219	+	65.00	131.78	319.90	539	AV432219 AV432219 Porphyra yezc
gb_est38:AV435766	+	65.00	131.62	326.43	549	AV435766 AV435766 Porphyra yezc
gb_est38:AV436271	+	65.00	131.59	327.74	551	AV436271 AV436271 Porphyra yezc
gb_est38:BG320355	+	65.00	128.02	518.35	836	BG320355 Zm03_05h01.F Zm03.AAFC
gb_est85:BF242509	+	65.00	127.53	551.85	885	BF242509 601876084F1 NIH-MGC-55
gb_est82:BF042036	+	64.50	130.89	358.47	529	BF042036 BP250023A20F6 Soares
gb_est74:BE403953	+	64.50	130.68	368.17	542	BE403953 WHE0415_C10_F1925 Whea
gb_est39:AV622430	+	64.50	130.64	370.41	545	AV622430 AV622430 Chlamydomonas
gb_est99:BG309910	-	64.50	127.10	582.80	823	BG309910 HVSMEC0014H07f Hordeum
gb_est78:BE731965	-	64.50	125.64	702.98	976	BE731965 601568983F1 NIH-MGC-21

gb_gss34:AZ922294 - 64.00 139.04 126.09 181 ! AZ922294 MRCot4H05 Sorghum b
gb_est96:BG063271 - 64.00 132.39 295.75 393 ! BG063271 H3005803-3 NIA Mous
gb_gss18:AQ883620 - 64.00 129.37 435.70 377 ! AQ883620 HS_5478_B2_B02_T7A
gb_est99:BG310261 - 64.00 125.79 689.85 849 ! BG310261 HVSMEC0016L07f Hord
seq_name: gb_est46:AW403229

seq_documentation_block:
LOCUS AW403229 439 bp mRNA EST 16-FEB-2000
DEFINITION UI-HF-BKO-abb-a-03-0-UI.r1 NIH_MGC_36 Homo sapiens cDNA clone
IMAGE:3055397 5', mRNA sequence.
ACCESSION AW403229
VERSION AW403229.1 GI:6922108
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 439)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov

Eco RI site shown at the beginning of the sequence.
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.
cDNA Library Preparation: M.B. Soares Lab
cDNA Library Arrayed by: M.B. Soares Lab
DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: M13 Forward.

FEATURES
Location/Qualifiers
1..439

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3055397"
/clone_lib="NIH_MGC_36"
/tissue_type="lymph"
/cell_type="germinal center B cells"
/cell_line="MGC85"
/lab_host="DH10B (LT1)"
/note="Vector: pMT3-Pac; Site_1: NotI; Site_2: Eco RI;
Constructed from size fractionated cytoplasmic mRNA
(0.5-1.5kb), directionally cloned. Cells provided by Louis
M. Staudt, Ph.D. Library preparation by Maria de Fatima
Bonaldo, Ph.D. and M. Bento Soares, Ph.D."

BASE COUNT 87 a 154 c 110 g 88 t

alignment_scores:
Quality: 178.00 Length: 32
Ratio: 5.562 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-485-951-1 x AW403229 ..

Align seg 1/1 to: AW403229 from: 1 to: 439

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17
|||||
232 AACCCCGGACAGTCCCTGTTGAGCTGCTTCTCCACGGTGGCTCTC 281
|||||
17 rGlnProValCysPheProProArgProArgGlyArgArgGlnLys 32
|||||
282 CCAGCCTGCTGTTTCCACCCAGGCCGCGGCGCAGACAAAA 327
|||||

seq_name: gb_est53:AW886539

seq_documentation_block:

REFERENCE
1 (bases 1 to 665)
AUTHORS
HCGP <http://www.ludwig.org.br/ORESTES>.
TITLE
The FAPESP/LICR Human Cancer Genome Project
JOURNAL
Unpublished (1999)
COMMENT
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3&t2=QV3-BT0537-221299-048-c06&t3=1999-12-22&t4=1>)
Seq primer: puc 18 forward
High quality sequence start: 23
High quality sequence stop: 664.

FEATURES
source
1..665
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="BT0537"
/dev_stage="Adult"
/note="Organ: breast; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the puc 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT 132 a 208 c 176 g 148 t 1 others
ORIGIN

alignment_scores
Quality: 164.00 Length: 32
Ratio: 5.467 Gaps: 0
Percent Similarity: 93.750 Percent Identity: 93.750
alignment_block:
US-09-485-951-1 x AW373817 ..

Align seg 1/1 to: AW373817 from: 1 to: 665

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17
|||||
485 AACCCCGGCACAGTCCCTGTCAGCCTGCTCTCCACGGTGGCTTC 534
17 rGlnProValCysPheProArgProArgGlyArgArgGlnLys 32
|||||
535 CCAGCCTGTGTGTTCCACCCAGGCGCGAGCGAGACACAA 580

seq_name: gb_est29:AL559254

seq_documentation_block:
LOCUS AL559254 535 bp mRNA EST 16-FEB-2001
DEFINITION AL559254 LTI_NFL008_Tc2 Homo sapiens cDNA clone CS0DJ012YD19 5
prime, mRNA sequence.
ACCESSION AL559254
VERSION AL559254.1 GI:12904575
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 535)
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope

Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CS0DJ012YD19"
/clone_lib="LTI_NFL008_Tc2"
/sex="male"
/tissue_type="T cells from T cell leukemia"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
<http://fulllength.invitrogen.com>"
BASE COUNT 97 a 183 c 126 g 89 t 40 others
ORIGIN

alignment_scores
Quality: 152.00 Length: 32
Ratio: 5.429 Gaps: 0
Percent Similarity: 87.500 Percent Identity: 87.500

alignment_block:
US-09-485-951-1 x AL559254 ..

Align seg 1/1 to: AL559254 from: 1 to: 535

1 AsnProArgThrValProValGlnProAlaPheSerThrValProPheSe 17
|||||
205 AACCCCGGCACAGTCCCTGTCAGCCTGCTCTCCACGGTGGCTTC 254
17 rGlnProValCysPheProArgProArgGlyArgArgGlnLys 32
|||||
255 CCAGCCTGTGTGTTCCACCCAGGCGCGAGCGAGACAA 300

seq_name: gb_est93:BF883419

seq_documentation_block:
LOCUS BF883419 404 bp mRNA EST 17-JAN-2001
DEFINITION QV3-ET0203-081200-525-a04 ET0203 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF883419
VERSION BF883419.1 GI:12273545
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE

AUTHORS

1 (bases 1 to 404)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

TITLE

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=QV3st2-QV3-ET0203-081200-525-804st3-2000-12-08&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 21
High quality sequence start: 404.

FEATURES

Location/Qualifiers
1..404
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="ET0203"
/dev_stage="Adult"
/note="Organ: lung_tumor; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT

80 a 150 c 87 g 87 t

ORIGIN

alignment_scores:

Quality: 118.00 Length: 22
Ratio: 5.364 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 90.909

alignment_block:

US-09-485-951-1 x BF883419 ..

Align seg 1/1 to: BF883419 from: 1 to: 404

7 ValGlnProAlaPheSerThrValProPheSerGlnProValCysPhePr 23

|||||

32 CTGGAGCCTGCTTCTCCAGCGTGGCTTCTCCAGCGCTGCTGTTCCTCC 81

23 oProArgProArgGly 28

|||||

82 ACCCAGGCCAGGGGC 97

seq_name: gb_est84:BF143767

seq_documentation_block:

LOCUS BF143767 703 bp mRNA EST 24-OCT-2000
DEFINITION 601789654F1 NCI_CGAP_Lu30 Mus musculus cDNA clone IMAGE:4020612 5',
mRNA sequence.

ACCESSION

BF143767

BF143767.1 GI:10982807

VERSION

EST.

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

NIH-MGC http://mgi.nci.nih.gov/

1 (bases 1 to 703)

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LLAM9274 row: k column: 13

High quality sequence stop: 685.

Location/Qualifiers

1..703

FEATURES

source

/organism="Mus musculus"
/strain="Czech II"
/db_xref="taxon:10090"
/clone="IMAGE:4020612"
/clone_lib="NCI_CGAP_Lu30"
/tissue_type="tumor, metastatic to mammary"
/lab_host="DH10B"
/note="Organ: lung; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SmaI; transgenic model WNT-1, expression driven by
MTV-LTR enhancer; Cloned unidirectionally. Primer: Oligo
dT. Library constructed by Life Technologies.

BASE COUNT

161 a 210 c 161 g 171 t

ORIGIN

alignment_scores:

Quality: 106.00 Length: 27
Ratio: 4.609 Gaps: 0
Percent Similarity: 85.185 Percent Identity: 74.074

alignment_block:

US-09-485-951-1 x BF143767 ..

Align seg 1/1 to: BF143767 from: 1 to: 703

6 ProValGlnProAlaPheSerThrValProPheSerGlnProValCysph 22

|||||

200 CTGTGCCAGCTGCTTCTCCAGCGTGGCTTCTCCAGCGTGTCTCTCAGCCAGTTCAGTT 249

22 eProProArgProArgGlyArgArgGlnLys 32

|||||

250 CCCACGGAGCCCTAAGGGCGCAACACAGAA 280

seq_name: gb_est84:BF181157

seq_documentation_block:

LOCUS BF181157 909 bp mRNA EST 31-OCT-2000
DEFINITION 601805820F1 NCI_CGAP_Mam5 Mus musculus cDNA clone IMAGE:4036661 5',
mRNA sequence.

ACCESSION BF181157

BF181157.1 GI:11059299

VERSION

EST.

KEYWORDS

EST.

SOURCE

house mouse.

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

NIH-MGC http://mgi.nci.nih.gov/

1 (bases 1 to 909)

National Institutes of Health, Mammalian Gene Collection (MGC)

Unpublished (1999)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Lohar Hennighausen Ph.D., Robin Humphreys

CDNA Library Preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

plate: LLAM9312 row: h column: 06

High quality sequence stop: 678.

Location/Qualifiers

1..909

FEATURES

source

/organism="Mus musculus"
/strain="C57/B6"
/db_xref="taxon:10090"
/clone="IMAGE:4036661"
/clone_lib="NCI_CGAP_Mam5"
/tissue_type="tumor, gross tissue"
/dev_stage="7 months"
/lab_host="DH10B"
/note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SmaI;

79 AACATCGCGCAGGTCCCAACGACGCCGCGCTGCCAAGGTGCAGTTC 128
17 rGlnProValCysPheProProArgProArgGlyArgArg 30
129 CCAGGCTGTGTGTCTCCACCCAGGCCAGGGGGGCGCAA 168

mRNA sequence.
 AA521753
 ACCESSION

REFERENCE	1 (bases 1 to 539)
AUTHORS	Smith,T.P.L., Casas,E., Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Laegreid, W.W. and Keefe,J.W.
TITLE	Design and use of four pooled tissue normalized cDNA libraries for EST discovery in cattle
JOURNAL	Unpublished (2000)
COMMENT	Contact: Smith rpl USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smith@email.marc.usda.gov Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 1 and -mismatch 12 options.

```

BASE COUNT
          142 a   156 c   154 g   138 t
/cldone110- barstead mouse proximal colon mpurbo
/dev_stage="7 day juvenile"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: EcoRI; Site_2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'.
TGTTACCAACTGAGTCGGAGCGGCCCTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[AAATCGATCCTTG], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pT73 vector.
Library constructed by Bob Barstead."

```

alignment_scores:		
Quality:	85.00	Length: 27
Ratio:	4.048	Gaps: 0
Percent Similarity:	77.778	Percent Identity: 62.963

alignment_block:
US-09-485-951-1 x AA521753

